

**MAPPING GRASSLAND DISTRIBUTION AND ABOVEGROUND BIOMASS IN
MANITOBA'S PRAIRIE ECOZONE USING MULTISENSOR SATELLITE IMAGES**

by

Mirmajid Mousavi

A Thesis

Submitted to the Faculty of Graduate Studies of

The University of Manitoba

In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department of Soil Science

University of Manitoba

Winnipeg, Manitoba

Copyright © 2025 by Mirmajid Mousavi

ABSTRACT

Mousavi, Mirmajid. M.Sci., The University of Manitoba, August 2025. Mapping Grassland Distribution and Aboveground Biomass in Manitoba's Prairie Ecozone Using Multisensor Satellite Images; Advisor: Nasem Badreldin.

Grasslands are one of the most widespread and endangered ecosystems in the world. They offer valuable ecosystem services, including food supply, carbon sequestration, soil fertility maintenance, biodiversity conservation, water and soil retention, and cultural benefits. Grasslands in the Canadian Prairies, the primary habitat for numerous grass species, have been increasingly degraded due to human interventions over the past decades. This has led to the loss of native species, environmental issues, and a decline in biodiversity. Therefore, monitoring the distribution of grassland and its biophysical characteristics, like aboveground biomass (AGB), is of the utmost importance for assessing ecosystem health and informing sustainable land management strategies. The overall main goals of this study were (1) to differentiate grasslands between native, tame and mixed classes using a novel artificial intelligence approach, and (2) to estimate grassland AGB in the Manitoba Prairie regions using various remote sensing datasets. For mapping grassland across the total Prairie regions of Manitoba, three advanced machine learning (ML) classification algorithms, including random forest (RF), gradient tree boosting (GTB) and support vector machine (SVM), were evaluated using input variables that came from different sources of remote sensing data, such as Landsat 8 and 9 (Landsat8-9), Sentinel-1 (S1), Sentinel-2 (S2), and Shuttle Radar Topography Mission (SRTM). Finding an optimal data combination (S1+S2) and using multi-temporal imagery were key components of satellite data processing aimed at improving grassland classification accuracy. Around 1,430 optical and radar images and 3,000 grassland samples were used to map grassland classes across the Manitoba Prairie Ecozone after excluding non-grassland classes, resulting in RF and GTB having higher performance with an overall accuracy of around 72%. The impact of multisensor image fusion on grassland classification accuracy was also assessed as an alternative approach when multi-temporal data composition was not feasible. For the second objective, parametric (e.g., linear and exponential regressions) and ML [RF, support vector regression (SVR) and artificial neural network (ANN)] regression methods were utilized to predict grassland AGB. The inclusion of environmental variables, such as temperature, precipitation and elevation, significantly improved the model

performance. Spectral data from three sensors with varying spatial resolution, including PlanetScope (3 m), S2 (10 m) and Landsat 8-9 (30 m), were used as model inputs to evaluate which dataset provided superior performance in biomass estimation. Model training and validation was based on newly derived biomass samples from daily Planet's Biomass Proxy products, which captured AGB dynamics at 10 m resolution across multiple growing seasons. This method enabled the observations of biomass to be extended throughout the growing season in multiple years. The findings showed that traditional parametric approaches ($R^2=0.63-0.65$) were lower than ML methods ($R^2=0.78-0.90$).

FORWARD

This thesis follows a manuscript-style format with the thesis guidelines set by the Department of Soil Science, University of Manitoba, and comprises two research manuscripts (Chapters 2 and 3). Each manuscript includes an abstract, introduction, materials and methods, results, discussion and conclusion. Chapter 1 provides a general introduction to grasslands, covering their types, ecological and economic benefits, the major threats they face, and an overview of the Prairie regions as Canada's primary grassland habitat. It also introduces the application of remote sensing and machine learning techniques for monitoring various aspects of grasslands, with further detail on the technologies and algorithms explored in this study. Chapter 4 is a comprehensive synthesis, summarizing the key findings and offering conclusions and recommendations.

Contributions of Authors:

This research was conceptualized and designed by Mirmajid Mousavi and Dr. Nasem Badreldin. Data were collected by AAFC and CFGA and analyzed by Mirmajid Mousavi and Nasem Badreldin. Two manuscripts were prepared by Mirmajid Mousavi and Dr. Nasem Badreldin. The first manuscript (Chapter 2) was published in the *Journal of Remote Sensing* (MDPI), while the second manuscript (part of Chapter 3) was published in the *IEEE International Symposium on Geoscience and Remote Sensing (IGARSS) 2025*.

The first manuscript was reviewed and edited by Mirmajid Mousavi, Dr. Nasem Badreldin, Dr. Marcos Cordeiro, Dr. James Kobina Mensah Biney, Dr. Barbara Kishchuk, Ali Youssef, Glen Friesen, Dr. Douglas Cattani, and Mustapha Namous. Data curation for this manuscript was conducted by Mirmajid Mousavi, Dr. Nasem Badreldin, Dr. Marcos Cordeiro, Dr. James Kobina Mensah Biney, and Ali Youssef. For the second manuscript, revision, editing, and data curation were undertaken by Mirmajid Mousavi and Dr. Nasem Badreldin.

ACKNOWLEDGEMENTS

“As you start to walk on the way, the way appears.”—Rumi, 13th-century Persian poet

I would like to thank my supervisor, Dr. Nasem Badreldin, with all my heart, as he supported, encouraged, and was very patient and generous to me during my studies. I am very happy that he trusted and believed in me to be able to accomplish this project. I also wish to express my sincere gratitude to Dr. Afua Adobea Mante, Dr. Marcos Cordeiro, and Dr. Douglas Cattani, who kindly accepted to serve as my advisory committee. They have been of great help; their advice and contributions have been of great value to this thesis.

I also want to thank several individuals from different organizations who assisted by supplying valuable field information or helping to facilitate the acquisition of data from their respective institutions. I want to thank Dr. James Biney (DAIab), Heather Pamm and Cedric MacLeod from the Canadian Forage and Grassland Association (CFGA) for the grassland classification ground data. Thanks to Mae Elsinger (Rangeland Biologist) from the Agricultural Ecosystems Unit in Agriculture and Agri-Food Canada (AAFC) for the biomass field survey data and Glenn Friesen from Manitoba Agriculture for facilitating the logistics of this research. Also, thanks to Dr. Cassidy Rankine and Austin Stone from Planet Labs PBC for the Biomass Proxy data support.

I am very grateful to my friends in Canada and Iran who supported me through difficult situations and never hesitated to help me when I needed it.

Finally, I express my deepest gratitude to my mother, my father, and my brothers for their unwavering support and boundless love I have received from them. Thank you for everything. This work is dedicated to the younger generation of my family—my dear nephews Mahbod, Amir Mahdi, Mohammad Taha, and Radin, and my lovely niece Noora—with the hope that it will inspire them in their future endeavors.

TABLE OF CONTENTS

ABSTRACT	ii
FORWARD	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xvi
1. INTRODUCTION	1
1.1. Grasslands, Services, and Threats	1
1.2. The Canadian Prairies: The Largest Habitat for Grasslands	6
1.3. The Landscape of Manitoba's PE.....	10
1.4. Grassland Mapping Using Remote Sensing Technology	14
1.5. The Development in Spaceborne Technology.....	18
1.5.1. Satellite Image Characteristics	20
1.6. The Application of Artificial Intelligence in Monitoring Grasslands.....	21
1.6.1. Decision Trees (DTs)	23
1.6.2. SVM.....	25
1.6.3. ANN.....	26
1.7. Thesis Objectives	28
1.8. Structure of Thesis.....	29
1.9. References	30
2. MAPPING GRASSLAND IN MANITOBA'S PRAIRIE ECOZONE.....	40
2.1. Abstract	40
2.2. Introduction.....	41
2.3. Materials and Methods	46

2.3.1. Study Area Description	46
2.3.2. Field Survey Sampling.....	47
2.3.3. Grassland Classification Strategy.....	49
2.3.4. Satellite Data Management.....	52
2.3.5. The Supervised Machine Learning Classification Models	56
2.3.6. Dimensionality Reduction Technique for Big Data	58
2.3.7. Classification Accuracy Assessment	60
2.3.8. Image Fusion.....	60
2.4. Results	62
2.4.1. Accuracy Assessment for Steps 1 and 2	62
2.4.2. Accuracy Assessment for Step 3.....	64
2.4.3. Accuracy Assessment for Sentinel-2+Landsat 8-9 Fused Data.....	68
2.4.4. The Detailed Grassland Map in Manitoba's PE	70
2.4.5. Landscape-Based Grassland Statistics.....	73
2.4.6. Implications to Policy Approaches for Sustainable Grassland Management	74
2.5. Discussion	75
2.6. Conclusions.....	78
2.7. Appendix.....	80
2.8. References	81
BRIDGE TO CHAPTER 3	91
3. GRASSLANDS ABOVEGROUND BIOMASS ESTIMATION USING REMOTE SENSING DATA IN MANITOBA'S PRAIRIE REGION	92
3.1. Abstract	92
3.2. Introduction.....	93
3.3. Materials and Methods	101

3.3.1. Study Area and Field Biomass Measurements	101
3.3.2. Grassland Biomass Prediction Strategy	104
3.3.3. Optical Remote Sensing Data	106
3.3.4. Biomass Proxy Raster Data	109
3.3.5. Landscape Topography Dataset	110
3.3.6. Climatic Data.....	110
3.3.7. Parametric Models	112
3.3.8. Machine Learning Models	114
3.3.9. Model Assessment	119
3.4. Results	120
3.4.1. Grassland AGB Estimation Using In-situ Samples	120
3.4.2. High-Temporal Biomass Prediction	125
3.4.3. Grassland AGB Estimation Using Synthetic Samples.....	128
3.5. Discussion	143
3.6. Conclusion	148
3.7. Appendix.....	150
3.8. References	152
4. OVERALL SYNTHESIS	162
4.1. Study Findings and Implications	162
4.2. Study Recommendations.....	164
4.3. References	165

LIST OF TABLES

Table 1.1. The list of several native and tame grasses found in the Manitoba Prairie Ecozone (PE), along with their seasonal growth types, was developed from Badreldin et al. (2021)	4
Table 2.1. The definition for grassland and other land use and land cover classes, adapted from Badreldin et al. (2021).....	48
Table 2.2. An overview of satellite datasets used in this study summarizes the satellite sensors employed in the study and highlights the acquisition dates, spatial resolution, and specific rationale for each sensor's inclusion in this study.....	52
Table 2.3. Description of all variables obtained through several sensors and utilized in this study.....	55
Table 2.4. Classification Accuracy assessment of step 1 (Vegetation vs. Non-Vegetation) and step 2 (Grassland vs. Non-Grassland) classifications using the RF method for different sensor combinations (PA% - UA%).	64
Table 2.5. The averaged OA and Kappa coefficient for different machine learning classification methods of the first two steps using Sentinel-1+Sentinel-2. The results show that DT-based classifiers have higher performance than SVM. However, RF achieved slightly higher accuracy than SVM in the first two steps.	64
Table 2.6. Accuracy assessment of machine learning methods with/without feature selection technique for grassland classification using Sentinel-1+Sentinel-2 data.	65
Table 2.7. Comparison of classification performance for fused images using BT and HSV fusion techniques, and non-fused images. The analysis included user's and producer's accuracy of tame, native and mixed classes.	69
Table 3.1: The details of grassland sampling sites, such as dates and grass types in PE.	104
Table 3.2. Spectral band characteristics of PlanetScope, Sentinel-2, and Landsat 8–9 OLI sensors used in this study; central wavelength in nanometers (nm), and spatial resolution for each sensor in meters (m), which is the size of the ground area represented by each pixel in the satellite image.	108
Table 3.3. Vegetation Indices (VIs) were generated using optical remote sensing images and applied in the research models.	109
Table 3.4. The coefficient of determination (R^2) of univariate linear and exponential regression models for biomass estimation using the selected vegetation indices (VIs)	121
Table 3.5. The comparison between parametric linear and exponential models using vegetation indices (VIs) with/without outlier samples	124
Table 3.6. The performance of linear regression with different combinations of predictors extracted from PlanetScope using synthesized biomass samples (n=668).	130
Table 3.7. The performance of linear regression with different combinations of predictors extracted from Sentinel-2 using synthesized biomass samples (n=628)	132

Table 3.8. The performance of linear regression with different combinations of predictors extracted from Landsat 8-9 using synthesized biomass samples (n=668)	133
Table 3.9. Coefficient estimates of linear regression models for PlanetScope, Sentinel-2 and Landsat 8-9 datasets, based on either all predictors or a selected subset.	134
Table 3.10. The performance of random forest (RF), support vector machine (SVR) and artificial neural network (ANN) models using various optical remote sensing sensors such as PlanetScope, Sentinel-2 and Landsat 8-9 in estimating grassland AGB.	136
Table 3.11. Principal components analysis (PCA) for the random forest (RF) and the linear regression models created using PlanetScope, Sentinel-2 and Landsat 8-9 data.....	137
Table 3.A1. The Pearson (r) and Spearman (ρ) correlations between biomass samples (n =102) and various vegetation indices (VIs) extracted from Sentinel-2 sensors.....	150

LIST OF FIGURES

Figure 1.1. Some of the native grasses are more common in Manitoba: from top to bottom and from left to right: (a) switchgrass, (b) prairie cordgrass, (c) nodding ryegrass, (d) needle-and-thread, (e) mannagrass, (f) little bluestem, (g) junegrass, (h) Indian-grass and (i) big bluestem. Photos were acquired by Dr. Douglas Cattani from the Department of Plant Science, University of Manitoba.	5
Figure 1.2. (a) The location of the Great Plains in North America. (b) The location of the Prairie Region within the Great Plains.	6
Figure 1.3. (a) Map of the Prairie Ecozone (PE) showing the PE boundary along with the locations of major cities, rivers, and lakes that are within or along the border of the PE. (b) The boundary of different ecoregions within PE.	7
Figure 1.4. The fluctuation of the annual average temperature in Canada between 1948 and 2023 (in degrees Celsius) from the 1961 to 1990 reference value (Bush et al., 2019; ECCC, 2025).	8
Figure 1.5. The various effects of climate change on grasslands in the Prairie Ecozone (PE). 10	
Figure 1.6. Prairie Ecozone (PE) in Manitoba and its ecoregions, including the Aspen Parkland, the Lake Manitoba Plain and the Southwest Manitoba Upland.	11
Figure 1.7. Digital elevation model (DEM) of Prairie Ecozone (PE) in the province of Manitoba, ranging from 218 to 760 meters. Elevation data were retrieved from SRTM with 30 m spatial resolution.	14
Figure 1.8. Remote sensing imagery of Winnipeg and neighbouring areas using (left) the Passive Sentinel-2 sensor and (right) the Active Sentinel-1 sensor.	19
Figure 1.9. Image contrasting lower (left) and higher (right) radiometric resolution, showing the greater level of grey scale (Canada Centre for Remote Sensing, 2010.).....	21
Figure 1.10. A complete decision tree (DT) model, where there is a hierarchical tree structure that begins with one main question at the top, called the root node. Internal decisions are new decision nodes that were leaf nodes before.	24
Figure 1.11. Graphical illustration of random forest (RF; top) and gradient tree boosting (GTB; bottom) algorithms (Jiao, Hu, et al., 2020; Jiao, Sun, et al., 2020).	25
Figure 1.12. Illustration of support vector machine algorithm. Left: there are several hyperplanes to separate blue balls from red balls, but only one optimal hyperplane with maximum margin. Right: optimal support vectors and margins.	26
Figure 1.13. Illustration of kernel separation in support vector machines (SVMs). In the left graph, a non-linearly separable dataset was shown before the kernel was applied. On the right, the same dataset after the kernel has been applied, and the classes become linearly separable using a hyperplane.	26
Figure 1.14. Illustration of an artificial neural network. The left figure shows the input and output structure of a biological neuron (Poddar, 2024). On the right, McCulloch-Pitts (or perceptron) provides a decision/output based on input data (Graupe, 2013). Its activation	

function and weighted inputs mimic the functions of biological axons and dendrites, respectively (Graupe, 2013).....	27
Figure 1.15. The architecture of an artificial neural network (ANN) model (Vawda et al., 2024).	28
Figure 2.1. Sankey diagram of land use and land cover (LULC) changes between 2012 and 2015 using annual crop inventory (ACI) data.	42
Figure 2.2. Geographical location of the study area. Manitoba’s PE with different ecoregions is located in Manitoba Province, Canada.	47
Figure 2.3. The spatial distribution of the ground-truthing sampling for all land cover and land use classes included in the classification of Manitoba’s PE grasslands.....	49
Figure 2.4. Workflow of the novel grassland classification strategy developed for enhanced ecological monitoring using multi-source remote sensing data and advanced machine learning techniques. The workflow integrates data from Sentinel-1, Sentinel-2, Landsat 8, and Landsat 9 satellites. It involves stages of image pre-processing, multi-temporal composition, image fusion using HSV and BT, and advanced machine learning classifiers, including RF, SVM, and GTB. The classification is performed in three steps to achieve fine-scale identification of native, tame, and mixed grasses, starting from basic vegetation classification (Level 1) to detailed grassland class differentiation (Level 3). Ancillary field data, topographic features, and LULC information were incorporated as inputs to generate the final grassland maps for web-based visualization.....	51
Figure 2.5. Atmospheric effects on remote sensing imagery demonstrate the influence of atmospheric conditions on the quality of remote sensing data, specifically how clouds and shadows can occlude pixels and impact the accuracy of the reflected signal received by MSS sensors; (a) Pixel occluded by shadow shows a scenario where a shadow, cast by an obstacle like a cloud, causes the pixel to be occluded, leading to distorted signals received by the sensor; and (b) Pixel occluded by cloud shows a situation where a cloud directly occludes the pixel, resulting in inaccurate data due to the cloud's interference with the reflected sunlight that reaches the sensor.	54
Figure 2.6. Multi-temporal mosaic of the study area during the growing season period using Sentinel-2 images.....	55
Figure 2.7. Median function used in Google Earth Engine (GEE) to reduce an image collection.	55
Figure 2.8. The steps of the HSV method, adapted from Al-Wassai et al. (Al-Wassai et al., 2011). After transforming the RGB image to HSV format, its V channel was replaced with the HR channel, which was then converted back to RGB mode.....	61
Figure 2.9. Comparison of step 2 grassland classification results using different machine learning models: (a) RF, (b) SVM, and (c) GTB.....	63
Figure 2.10. Producer’s Accuracy (PA) and User’s Accuracy (UA) of three machine learning models such as RF, GTB and SVM, with and without feature selection (FS).....	65

Figure 2.11. Classification accuracy varies with different input features ranked based on ANOVA for the RF, SVM, and GTB.	66
Figure 2.12. Out-of-bag (OOB) error for different numbers of trees and the number of variables (<i>mtry</i>) per split was calculated. Different <i>mtry</i> per split tested are the square root of the total number of variables (SQRT), the total number of variables (ALL), and the natural logarithm of the total number of variables (Ln).	67
Figure 2.13. (a) Overall accuracy (OA) of classification for different kernel types in SVM model. (b) Grid search to find the best value for the Cost/Regularization parameter for Linear kernel in SVM.	67
Figure 2.14. The effect of the number of trees (<i>ntree</i>) on OA in GTB classification	67
Figure 2.15. Subsetted areas before and after image fusion for image quality improvement; (a) Landsat 30 m, (b) HSV fused image, (c) BT fused image.	68
Figure 2.16. Scatter plots of fused and non-fused bands using BT and HSV approaches. Except for band 2 (B2), HSV had a higher R-squared. Around 1900 points were selected to build the scatter plots, and the colour bar represents the point density, ranging from low density (Blue) to high density (Red).	69
Figure 2.17. The classification maps of pixel-level fusion with RF approach using (a) the multispectral image, (b) the HSV fused image, and (c) the BT fused image.	70
Figure 2.18. The detailed grassland classification of Manitoba's PE using the RF classification machine learning model and Sentinel-1 and Sentinel-2 data combination.	72
Figure 2.19. Distribution of mixed, tamed, and native grasslands across three ecoregions, highlighting the areas covered by each grassland class. The percentage listed for each ecoregion shows its proportion of the total grassland area, with Southwest Manitoba Uplands at 2.42%, Lake Manitoba Plain at 41.83%, and Aspen Parkland at 55.75%. The relative dominance of each grassland type across Aspen Parkland, Lake Manitoba Plain, and Southwest Manitoba Uplands illustrates regional differences in land use and ecological composition.	73
Figure 2.A1. The list of all features with their scores. The numbers following spectral bands, VIs and backscatter variables indicate multiple composite images created during the growing season. Red points show the features that were excluded from classification models to achieve their highest OA and Kappa coefficient; (a) ANOVA F-Value of RF; (b) ANOVA F-Value of SVM; and (c) ANOVA F-Value of GTB.	80
Figure 3.1. The rising plate meter (RPM) is a common tool for grassland AGB estimation (Rankin & Komainda, 2023).	94
Figure 3.2. Map of the study area in Manitoba, Canada, used for biomass and biodiversity modelling. (a) Location of the Prairie Ecozone (PE) within the province of Manitoba. (b) Distribution of ecoregions within Manitoba's PE, illustrating the spatial variability of ecological zones across the study area.	102

Figure 3.3. Location of six fields where biomass was surveyed in Manitoba's PE (red rectangles) and distribution of biomass samples at each field. Field's name was embedded in them.....	103
Figure 3.4. General workflow of the primary steps in the grassland AGB estimation strategy. Step 1 was preparing datasets. Vegetation indices (VIs) extracted from Sentinel-2 and Biomass Proxy were acquired according to the timing of ground-truth data collection. In step 2, the fit of several univariate parametric models in AGB prediction was assessed, and Biomass Proxy which offers daily products, was used to generate new biomass samples for a more thorough evaluation on biomass. Step 3 involved a comprehensive evaluation using synthetic biomass samples with multivariate parametric regression models and machine learning algorithms, incorporating data from three optical satellite sensors (PlanetScope, Sentinel-2, and Landsat 8-9). At this step, the impact of environmental data like precipitation, temperature and elevation was also analyzed.....	105
Figure 3.5. Illustrates the processing chain and steps involved to generate each of PlanetScope's imagery products (PlanetScope Planet Documentation, 2025).....	107
Figure 3.6. Monthly precipitation temperature variation over the 30-year period from 1993 to 2023. Each precipitation box represents the distribution of total precipitation for a specific month across all years, while each temperature box shows the distribution of the monthly average temperature, calculated as the mean of daily maximum and minimum temperatures over the same period. Panels (a) and (d), (b) and (e), and (c) and (f) represent the northern, central, and southern regions of the Prairie regions of Manitoba, respectively.	112
Figure 3.7. Random forest (RF) components in the regression task during the training process. K bootstrap samples are produced using the training data, whereby, for each sample generated, an individual decision tree is built on a random subset of the features. Two-thirds (In-Bag) of each bootstrap sample is used for training and one-third (OOB) is used to test internal validation.....	116
Figure 3.8. The support vector regression (SVR) with ϵ -tube. The central line represents the regression function, while the dashed line defines the ϵ -insensitive margin (ϵ -tube). If data points are placed inside the tube, they are not penalized during training.	117
Figure 3.9. The ReLU activation function and its plot. Despite its simple form, ReLU's non-linearity allows it to learn more complex data patterns between features.	118
Figure 3.10. Histogram distribution of AGB across validated grassland sampling points, overlaid with a kernel density estimation (KDE) curve.....	120
Figure 3.11. Diagnostic plot of leverage and Cook's Distance for the linear regression model using the RVI predictor ($n = 102$). The figure shows the scatter plot of Cook's Distance versus leverage, with the common influence threshold ($4/n$) indicated by a dashed red line. Observations classified as influential (Cook's $D > 4/n$) are highlighted in red and labelled by their observation index. Boxplots of Cook's Distance values are shown along both graph axes.	123

Figure 3.12. Coefficient of determination (R^2) for linear and exponential regression models between Biomass Proxy and in-situ data at different temporal offsets with/without outliers considered.....	125
Figure 3.13. Temporal variation of synthesized biomass values during the growing seasons from 2019 to 2023 at the locations of all biomass samples across six fields (the site ID at the top left of each plot corresponds to the geographic position of the site as shown in Figure 3.3), produced using the Biomass Proxy predictor. Daily precipitation is also presented in the second Y-axis.....	128
Figure 3.14. Pairwise Pearson correlation (r) between grassland AGB with spectral bands of PlanetScope, vegetation indices (VIs), elevation (DEM), precipitation (Prcp) and temperature (Temp).....	129
Figure 3.15. Pairwise Pearson correlation (r) between grassland AGB with spectral bands of Sentinel-2, vegetation indices (VIs), elevation (DEM), precipitation (Prcp) and temperature (Temp).....	131
Figure 3.16. Pairwise Pearson correlation (r) between grassland AGB with spectral bands of Landsat 8-9, vegetation indices (VIs), elevation (DEM), precipitation (Prcp) and temperature (Temp).....	133
Figure 3.17. Predictor importance of modelling of grassland AGB across six fields in the Prairie Ecozone (PE). (a) Importance of nine driving variables individually and (b) within categories	138
Figure 3.18. Relationship between observed and estimated AGB using parametric (linear and exponential) models based on PlanetScope, Sentinel-2 and Landsat 8-9 optical data: (a) linear-PlanetScope, (b) exponential-PlanetScope, (c) linear-Sentinel-2, (d) exponential-Sentinel-2, (e) linear-Landsat 8-9, and (f) exponential-Landsat 8-9	138
Figure 3.19. Relationship between observed and estimated AGB using machine learning [random forest (RF), support vector regression (SVR) and artificial neural network (ANN)] models based on PlanetScope, Sentinel-2 and Landsat 8-9 optical data: (a) RF-PlanetScope, (b) SVR-PlanetScope, (c) ANN-PlanetScope, (d) RF-Sentinel-2, (e) SVR-Sentinel-2, (f) ANN-Sentinel-2, (d) RF-Landsat 8-9, (e) SVR-Landsat 8-9, and (f) ANN-Landsat 8-9	139
Figure 3.20. Maps of grassland AGB estimated using the random forest (RF) model with input data from (a) PlanetScope, (b) Sentinel-2, and (c) Landsat 8-9, acquired at different locations and dates.....	142
Figure 3.A1. Residual distribution of predicted versus ground-estimated AGB for six models (a-f), corresponding to the linear-PlanetScope, exponential-PlanetScope, linear-Sentinel-2, exponential-Sentinel-2, linear-Landsat 8-9, and exponential-Landsat 8-9.....	151
Figure 3.A2. Residual distribution of predicted versus ground-estimated AGB for six models (a-i), corresponding to the RF-PlanetScope, SVR-PlanetScope, ANN-PlanetScope, RF-Sentinel-2, SVR-Sentinel-2, ANN-Sentinel-2, RF-Landsat 8-9, SVR-Landsat 8-9, and ANN-Landsat 8-9	151

LIST OF ABBREVIATIONS

ACI	Annual Crop Inventory
AGB	Aboveground Biomass
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BOA	Bottom of Atmosphere
BP	Biomass Proxy
BT	Brovey Transform
CFGF	Canadian Forage and Grassland Association
CMI	Climate Moisture Index
CNNs	Convolutional Neural Networks
CR	Cross-Ratio
CV	Cross Validation
dB	Decibels
DD	Degree Days
DEM	Digital Elevation Model
DM	Dry Matter
DNN	Deep Neural Network
DTs	Decision Trees
ESA	European Space Agency
EVI	Enhanced Vegetation Index
GEE	Google Earth Engine
GI	Gini Index
GLA	Green Leaf Algorithm
GNDVI	Normalized Difference Vegetation Index
GPP	Gross Primary Productivity
GRD	Ground Range Detected
GTB	Gradient Tree Boosting
HR	High-Resolution
HRVIR	High Resolution Visible and Infrared
HSV	Hue, Saturation, and Value
IFOV	Instantaneous Field of View
LAI	Leaf Area Index
LiDAR	Light Detection and Ranging
LULC	Land Use and Land Cover
m	meters
MGI	Manitoba Grassland Inventory
ML	Machine Learning
MODIS	Moderate Resolution Imaging Spectrometer
MSI	Multispectral Image
MSS	Multispectral Spaceborne Sensors
mtry	Number of Features
NASA	National Aeronautics and Space Administration
NPP	Net Primary Productivity

NIR	Near-infrared
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
ntree	Number of Trees
OA	Overall Accuracy
OLI	Operational Land Imager
OOB	Out-of-Bag
PA	Producer's Accuracy
PCA	Principal Component Analysis
PE	Prairie Ecozone
PSRI	Plant Senescence Reflectance Index
PanNDVI	Panchromatic Normalized Difference Vegetation Index
Prcp	Precipitation
QA	Quality Assurance
R ²	Coefficient of Determination
RADAR	Radio Detection and Ranging
RBF	Radial Basis Function
RDVI	Renormalized Difference Vegetation Index
RF	Random Forest
RMSE	Root Mean Square Error
RVI	Ratio Vegetation Index
SAR	Synthetic Aperture Radar
SOC	Soil Organic Carbon
SLC	Soil Landscapes of Canada
SRTM	Shuttle Radar Topography Mission
SWIR	Shortwave Infrared
SVM	Support Vector Machine
SVR	Support Vector Regression
TOA	Top of Atmosphere
Temp	Temperature
UA	User's Accuracy
UAV	Unmanned Aerial Vehicle
VIs	Vegetation Indices
r	Pearson Correlation
ρ	Spearman Correlation

1. INTRODUCTION

1.1. Grasslands, Services, and Threats

Grasslands account for around 40% of the Earth's terrestrial surface (Y. Wu et al., 2023). Biologically, grasslands are one of the most diverse ecosystems, particularly in terms of species at risk (Roch & Jaeger, 2014). Various scholars have defined grassland in many specific ways. Milner (1968) defined it as “a plant community in which Gramineae dominates and trees are absent”; Bazzaz & Parrish (1982) specified grassland as “a vegetation type dominated by grasses but containing many broadleaf herbs (forbs)”; and Knapp et al. (1998) described it as “a general lack of woody vegetation helps define grassland.” A common feature observed in all various definitions is a low abundance of woody vegetation.

Grasslands are not a homogeneous landscape; they are highly diverse and vary greatly across different soil zones and climates (Blair et al., 2014). There are several major types of grasslands globally, including temperate grasslands such as the Prairies of North America and the steppes of Eurasia, tropical savannas which are found in Africa, South America, and parts of Australia, and high-altitude grasslands like those ecosystems in the Andes and the Himalayas (Blair et al., 2014). Each of these grassland ecosystems has distinct climatic conditions, species compositions, and ecological processes. For instance, tropical savannas often include a mixture of grasses and scattered trees, whereas temperate grasslands like prairies are typically treeless and experience more noticeable seasonal changes (Blair et al., 2014).

Soils in grassland regions tend to be rich in organic matter due to the continuous cycle of plant growth and decay (Blair et al., 2014). Most of these soils, such as the Chernozem (i.e., Mollisols according to the US Soil Taxonomy), which is found in North American Prairies, are among the most fertile soils in the world, making them ideal for agriculture (Bai & Cotrufo, 2022). However, the conversion of native grasslands to cropland has significantly reduced their extent and disrupted the ecological balance (Lark et al., 2020). This transformation often leads to habitat loss, reduced biodiversity, and alterations in water and nutrient cycles (Blair et al., 2014), and impacts the primary ecosystem's services provided by grasslands, such as carbon sequestration and ecological integrity (Cai et al., 2024; Dao et al., 2021). In comparison to forests, the carbon pool in grasslands is smaller; however, they represent approximately

25-34% of the global terrestrial carbon store (Liu et al. 2023). The effectiveness of grasslands in carbon storage is greater for underground tissues compared to aboveground ones; This higher-level below-ground productivity is attributed to the extensive root systems, while aboveground production accounts for less than 30% of the net primary production (Cao et al., 2019). Grasslands are essential ecosystems that significantly contribute to the agricultural sector by providing a sustainable source of nutritious forage for livestock, thereby enhancing food security and supporting livestock health and productivity. (Bengtsson et al., 2019). In Canada, roughly one-third of agricultural lands (equivalent to 21.1 million hectares, Mha) are dedicated to the beef sector (Pogue et al., 2018). This includes 12.9 Mha of native grassland and 5 Mha of tame or cultivated grassland (Pogue et al., 2018).

Grasslands are vital carbon sinks, as they are capable of storing a large amount of carbon in both their above- and below-ground structures. The largest portion of their carbon is stored underground in the form of soil organic carbon (SOC) (Ghahramani et al., 2019). The main sources of this carbon are plant roots, root exudates, and microorganisms, which contribute to improved soil fertility, ecosystem stability, and climate regulation. Native grasslands can store a larger amount of carbon because of their deeper root systems (Ghahramani et al., 2019). Climate change is expected to affect the balance between carbon inputs and outputs in grassland soils due to changes in temperature, precipitation, and plant composition. SOC in grasslands is stabilized as particulate organic matter (POM) and mineral-associated organic matter (MAOM) (Bai & Cotrufo, 2022). Dass et al. (2018) reported that in the 21st century, rangelands and grasslands function better as carbon sinks than forests since they are better able to withstand drought and wildfire. Forests contain most of their carbon in trees, which is exposed to the air if fires happen or climate change strikes; in contrast, grasslands move carbon to underground storage, protecting it during fires. They experimented with different model simulations under various climate scenarios, confirming that grasslands absorb more carbon dioxide than forests, except when global emission cuts are substantial (Dass et al., 2018).

Many grassland species, including native and tame types, are found in the Prairie region, which is a primary habitat for grasslands (National Grassland Inventory, 2024). Most native grasses were degraded in the Prairie due to human disturbance, notably by expanding agriculture over the past decades (Scott, 1996). Native grasslands are rich in species diversity

and support ecological processes such as pollination and soil and water regulation (Cook, 2019; Petermann & Buzhdygan, 2021). The significant loss of native grasses has affected the dependent species. For instance, grassland birds experienced about a 67% population decline since 1970, with several species, like the Chestnut-collared Longspur and Sprague's Pipit, facing the risk of complete extinction (Cook, 2019). Moreover, grasslands help prevent soil erosion by preserving soil cohesion and forming a tightly packed root system beneath (Watersheds, 2023). The root system enhances soil structure by creating channels for air and water movement (Srivastava & Yetgin, 2024), and native grasses have deeper root systems, allowing them to sequester more carbon than cultivated grasses (Crews & Rumsey, 2017). The natural grasslands of Manitoba can be classified into four major groups: Tallgrass (south-central Manitoba), Mixed-grass (southwestern Manitoba), Sandhill (southwestern and south-central Manitoba) and Fescue (found in Aspen Parkland) (Prairie Grasslands, 2014). The main native grasslands found in Manitoba may include big bluestem, Indian-grass, western wheat-grass, rough fescue and little bluestem (Forage Species, 2024).

Native grasslands have been facing serious competition from non-native grasses that humans introduced for agricultural goals (Zapisocki et al., 2022). These non-native grasses have emerged as invasive species in the native Prairie ecosystem and have been replacing native grasses (Zapisocki et al., 2022). For example, Kentucky bluegrass (or bluegrass), which now spans the entire Prairie, is a perennial species of grass native to all of Europe and North Asia (Carrier & Bort, 1916). The grasses that expanded primarily for agricultural purposes, such as livestock grazing, may be considered tame or seeded grasses. In fact, tame grasses are those cultivated as forage for cattle and other grazing animals (Ren et al., 2018). Manitoba is a major centre for forage and turf seed production in Canada (Agriculture | Province of Manitoba, 2024). Various types of grasses and legumes, such as perennial ryegrass, tall fescue, timothy, Kentucky bluegrass, and alfalfa, are grown in Manitoba. Some grasses, like Kentucky bluegrass and fine fescue, are widely used as turf grasses for recreational activities (Agriculture | Province of Manitoba, 2024).

Grasses are also classified into C3 (cool-season) or C4 (warm-season) depending on the type of photosynthesis they use and the attributes of their leaves (Belesky & Fedders, 1995). These differences affect the species' adaptations to various environmental conditions as well as the nutritional value and environmental resources used by the plants (Forage Species,

2024). Most tame grasses found in Canada are C3 species and usually thrive best when the temperature ranges from 18°C to 24°C. They thrive in cool weather and experience their peak growth in the spring and early summer months and may go dormant under warmer, drier conditions. Perennial C3 grasses include brome grass, fescues, ryegrass and timothy. Native C3 grasses include slender wheat-grass, plains rough fescue, needle-and-thread and june-grass (Forage Species, 2024). C3 forages tend to produce higher protein content compared to C4 grasses and provide the majority of their annual biomass by mid-July. On the other hand, C4 grasses can tolerate hot, dry conditions and include corn, sorghum, sorghum-sudan-grass, sudan-grass and millet, as well as native grasses such as big and little bluestem, side-oats grama, Indian-grass and blue grama. These species thrive in summer, with peak productivity occurring from late June through August, making them valuable for late summer or fall grazing (Forage Species, 2024). The list of popular native and tame grasses located in Manitoba province is provided in Table 1.1, and the pictures of some native grasses are illustrated in Figure 1.1.

Table 1.1. The list of several native and tame grasses found in the Manitoba Prairie Ecozone (PE), along with their seasonal growth types, was developed from Badreldin et al. (2021)

Grass Type	Seasonal Type	Grass Name
Native	Warm/Cool-seasons	▪ Little Bluestem, (<i>Schizachyrium scoparium</i>) ▪ Switchgrass, (<i>Panicum virgatum</i>) ▪ Indian-Grass, (<i>Sorghastrum nutans</i>) ▪ Blue Grama, (<i>Bouteloua gracilis</i>) ▪ Side-oats Grama, (<i>Bouteloua curtipendula</i>) ▪ Needle-and-thread Grass, (<i>Hesperostipa comata</i>) ▪ June Grass, (<i>Koeleria macrantha</i>) ▪ Plains Rough Fescue, (<i>Festuca hallii</i>) ▪ Slender Wheat-grass, (<i>Elymus trachycaulus</i>) ▪ Green Needle Grass, (<i>Stipa viridula</i>) ▪ Canada Wild Rye, (<i>Elymus canadensis</i>) ▪ Giant Wild Rye, (<i>Elymus cinereus</i>) ▪ Prairie Dropseed, (<i>Sporobolus heterolepis</i>) ▪ Tufted Hairgrass, (<i>Deschampsia caespitosa</i>) ▪ Sweetgrass, (<i>Hierochloa odorata</i>) ▪ Sheep Fescue, (<i>Festuca ovina saximontana</i>) ▪ Purple Oat Grass, (<i>Schizachne purpurascens</i>) ▪ Marsh Reed Grass, (<i>Calamagrostis canadensis</i>) ▪ Tall Manna Grass, (<i>Glyceria grandis</i>) ▪ Buffalo Grass, (<i>Buchloe dactyloides</i>)
Tame	Warm/Cool-season	▪ Smooth Brome grass, (<i>Bromus inermis</i>) ▪ Meadow Brome grass, (<i>Bromus riparius</i>) ▪ Timothy, (<i>Phleum pratense</i>) ▪ Orchardgrass, (<i>Dactylis glomerata</i>) ▪ Tall Fescue, (<i>Festuca arundinacea</i>) ▪ Crested Wheat-grass, (<i>Agropyron cristatum</i>)

		▪ Perennial Ryegrass, (<i>Lolium perenne</i>) ▪ Intermediate Wheat-grass, (<i>Thinopyrum intermedium</i>) ▪ Kentucky Bluegrass, (<i>Poa pratensis</i>) ▪ Reed Canary Grass, (<i>Phalaris arundinacea</i>) ▪ Russian Wild Ryegrass (<i>Psathyrostachys juncea</i>) ▪ Sudan-grass, (<i>Sorghum × drummondii</i>)
--	--	---

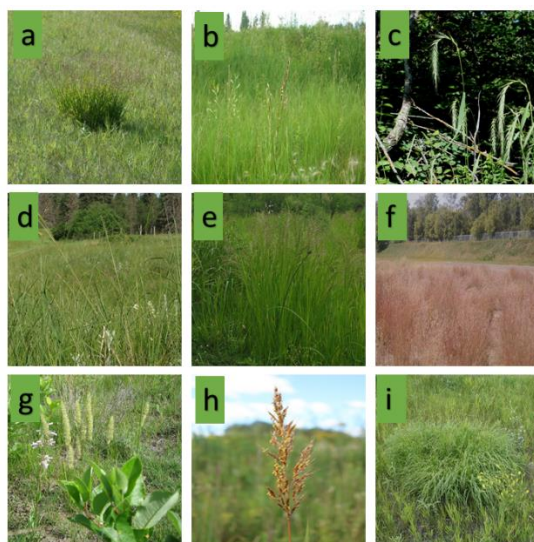


Figure 1.1. Some of the native grasses are more common in Manitoba: from top to bottom and from left to right: (a) switchgrass, (b) prairie cordgrass, (c) nodding ryegrass, (d) needle-and-thread, (e) mannagrass, (f) little bluestem, (g) junegrass, (h) Indian-grass and (i) big bluestem. Photos were acquired by Dr. Douglas Cattani from the Department of Plant Science, University of Manitoba.

Many grassland ecosystems worldwide are particularly vulnerable to climate shifts, and the Canadian Prairie provinces, such as Manitoba, are already experiencing the impacts (Bush et al., 2019). For example, the Prairies experienced an increase in average annual temperatures of 1.5 to 2.0°C during the twentieth century and are expected to continue warming in the future (Bush et al., 2019; Masson-Delmotte et al., 2021). This increase causes the lengthening of the growing season, which initially enhances biomass productivity for certain grass species. However, these early-season increases are commonly outweighed by heat stress and rapid phenological development in C3 species (Tilman et al., 2006). Advanced green-up and flowering lead to a decrease in forage quality and may restrict regrowth during the important summer grazing season (Tilman et al., 2006). C4 grasses are likely to thrive and become dominant under warmer conditions in mixed grasslands (Havrilla et al., 2023). This shift in ecological dominance might result in an increased demand for forage during the late summer and fall periods (Vendramini et al., 2023).

1.2. The Canadian Prairies: The Largest Habitat for Grasslands

The Great Plains is one of the ecological regions of North America, which spans over three countries, including three Mexican states, 18 U.S. states and three Canadian provinces (Figure 1.2). The northern part of the Great Plains, which accounts for approximately 16% of the region, is located in Canada. This region, which looks like a triangle, is the Prairie Ecozone (PE), which makes up around 5% of Canada's landmass (Shorthouse, 2010).

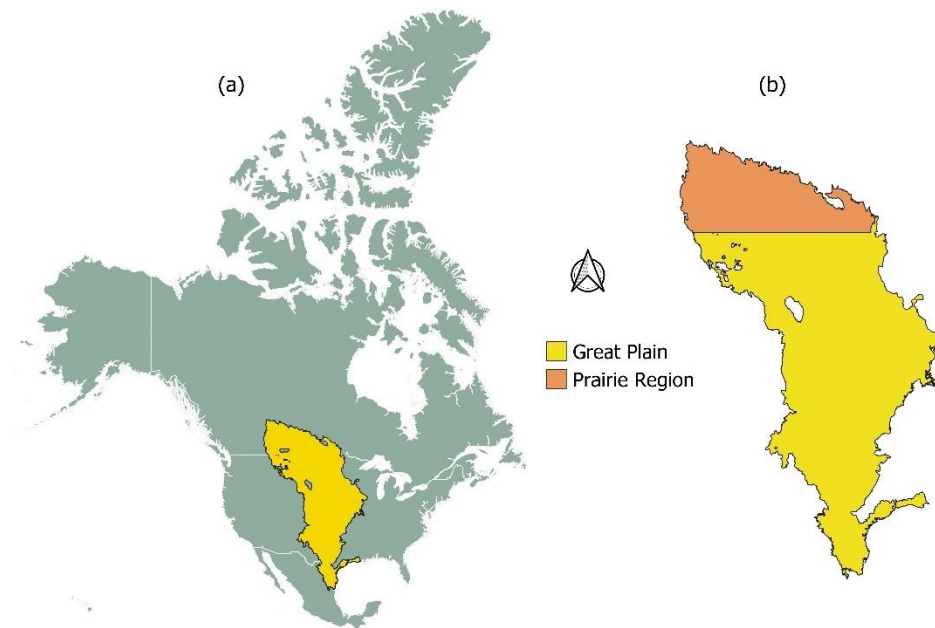


Figure 1.2. (a) The location of the Great Plains in North America. (b) The location of the Prairie Region within the Great Plains.

As shown in Figure 1.3, the Canadian PE is located in three provinces of Alberta, Saskatchewan and Manitoba, and it is subdivided into seven ecoregions: Aspen Parkland, Cypress Grassland, Fescue Grassland, Lake Manitoba Plain, Mixed Grassland, Moist Mixed Grassland and Southwest Manitoba Upland (Shorthouse, 2010). Aspen Parkland is the only ecoregion that exists across all Prairie provinces. The PE has a semi-arid to sub-humid climate, supporting extensive and temperate grasslands (Government of Canada, 2002). The main land cover in the PE before European settlement was native grasslands. However, agricultural expansion has transformed cropland into the primary land use in the PE (Government of Canada, 2002). Thus, many services provided by grasslands, including wildlife conservation, soil carbon sequestration and biodiversity maintenance (Badreldin et al., 2021), have decreased over the past two centuries (Government of Canada, 2002). The latest report

estimated that only 25-30% of native grasslands were preserved in the PE (Roch & Jaeger, 2014).

People have lived in the Prairies for 10,000 years, and the A'aninin, Assiniboine, Cree, Sioux, and Blackfoot First Nations would often follow the migrating bison as they travelled in grassland areas. This long history of connection to the Prairies is one indicator of how important grasslands have been in First Nation's culture. Grassland habitats supplied plants and animals for humans as medicine, food, materials, a home, and a source of culture, while Indigenous peoples helped to preserve and care for them (Whitlaw, 2022).

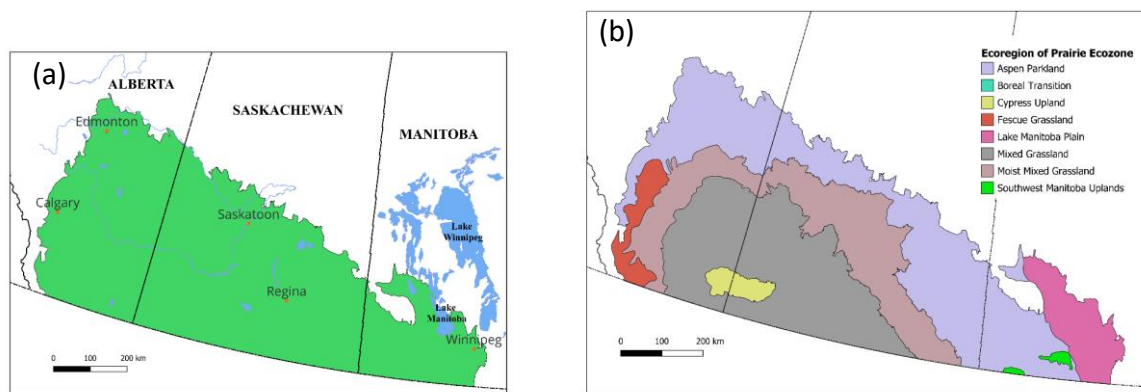


Figure 1.3. (a) Map of the Prairie Ecozone (PE) showing the PE boundary along with the locations of major cities, rivers, and lakes that are within or along the border of the PE. (b) The boundary of different ecoregions within PE.

Annual rainfall in these grasslands is very variable, about 350 mm in some years and as much as 500 or 550 mm in others (McGinn & Shepherd, 2003). Around 70–80% of the annual rainfall in the Prairies is received mainly as rain, with June and July together making up 20–35% of the yearly total rainfall, and the growing season comprises 42–54% of that. For the southern grasslands, the peak month for rainfall is June, while at higher latitudes, the peak month for precipitation is July (Chakravarti, 1972; Wheaton, 1998).

Regarding the average annual maximum temperature ($T_{\max}$), in Alberta, Saskatchewan and Manitoba is 8.4°C, 8.3°C and 7.6°C, respectively. Winter experiences $T_{\max}$ measuring around -2.5°C in southwestern Alberta due to the chinook-driven warming along the Rocky Mountains, but average temperatures plummet below -10.5°C in other provinces because of the cold continental air. The average maximum temperature in summer in the region is about

25.5°C in the south, but it decreases with both latitude and altitude (S. McGinn & Shepherd, 2003).

The warming in the agricultural region of Prairie since 1951 has been reported by Cutforth and Judiesch (2007), with annual average daily maximum and minimum air temperatures increasing by 0.2 and 0.3 °C per decade, respectively, between 1951 and 2004. The meteorological data from 1950 to 2004 also indicated that January-April experienced the largest air temperature increase in the Prairie, with maximum and minimum air temperatures increasing by 4.9 and 4.8 °C, respectively (Cutforth & Judiesch 2012).

Knowing about climate change helps explain the changing conditions of prairie habitats. In this area, one of the main changes from climate change is a continuous increase in temperature. The data from 1900 to 1998 show that the number of very cold days in winter, spring and summer dropped in southern Canada and that extremely warm days became more frequent in winter and spring (Bonsal et al., 2001). In addition, from 1950 up to 1989, the average recorded maximum and minimum temperatures increased by 1.7°C and 1.1°C (Figure 1.4), respectively (Skinner & Gullett, 1993).

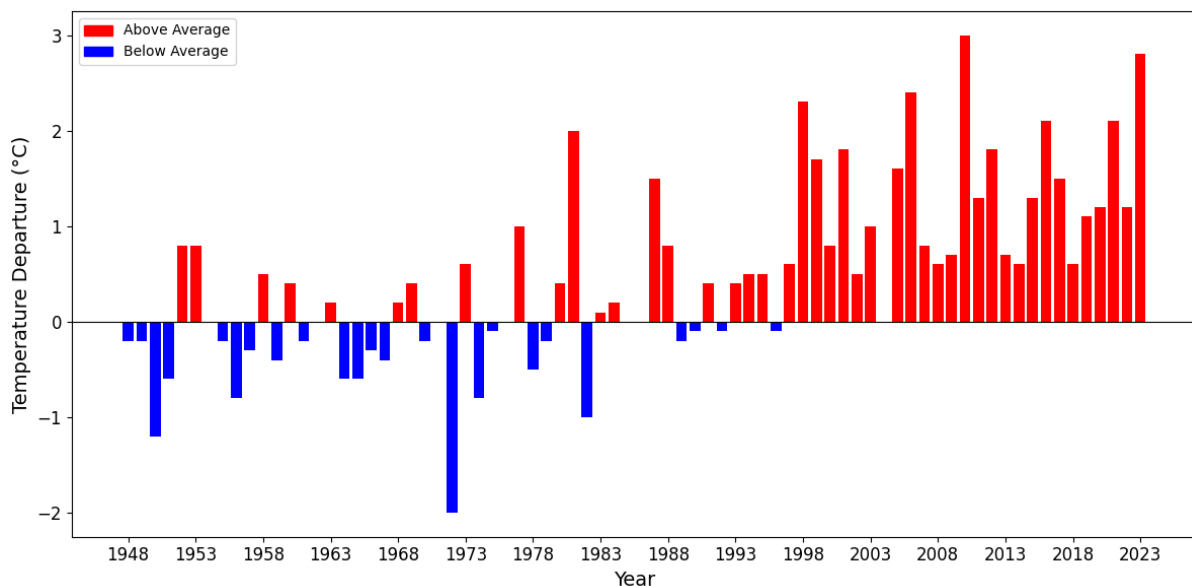


Figure 1.4. The fluctuation of the annual average temperature in Canada between 1948 and 2023 (in degrees Celsius) from the 1961 to 1990 reference value (Bush et al., 2019; ECCC, 2025).

Apart from being warmer, the prairie region has also experienced different precipitation patterns. Between 1951 and 2004, the incidence of rainfall events (i.e., any day that rain fell) over the Prairies had risen by 4 per decade, but the intensity of those rainfall

events had fallen by 4.4% per decade (Mapfumo et al., 2023). Rainfall has risen by about 16% from 1956 to 1995, probably due to climate change making spring warmer and earlier, which causes snowfall to become rainfall (Mapfumo et al., 2023).

The current climate projections predict more frequent extreme rain or drought conditions in the future (Climate Atlas of Canada, 2022), such as drought stress during early summer, which significantly reduces the performance of shallow-rooted C3 grasses and legumes. Some native prairie grasses, like big bluestem and switchgrass, may be better able to withstand changing precipitation patterns due to their deep root systems (Tilman & El Haddi, 1992). Spring rainfall plays a key role in the successful growth and yield of C3 pasture crops, while sufficient rain during mid-to-late summer benefits the development of C4 grasses (McGinn, 2010). Climate variability makes it challenging to pasture stewardship activities such as grazing, seed production and hay harvesting (Craine et al., 2012). The changing climatic conditions contribute to plant species composition changes in both native and tame grassland (Wangchuk et al., 2021). Higher temperatures and more frequent disruptions, such as drought and overgrazing, increasingly enable invasive plant species, like Kentucky bluegrass and crested wheat-grass, to become dominant, thus decreasing the plant diversity (Liu et al., 2018). Such shifts in community structure may impair the functionality of ecosystems, change soil microbial communities, and hence result in the local extinction of native species (Liu et al., 2018; Schüle et al., 2023). The increase in temperatures could also decrease the nutrient density of forage (Lee et al., 2017). C3 grasses tend to have better forage quality in the early part of the growing season, while C4 grasses, which are more resistant to drought, often have lower protein and higher lignin levels (Forage Species, 2024). Grazing management practices such as shifting grazing times, planting tolerant grasses, or increasing soil fertility can help mitigate the effects of climate disruptions on forage nutrient levels, see Figure 1.5 (Ghahramani et al., 2019).

Warming can both increase and decrease SOC depending on context-specific conditions, reflecting the heterogeneous response of grassland soils to climate change. In semi-arid regions, where microbial activity is moisture-limited, warming may stimulate microbial processes and enhance carbon stabilization in MAOM through increased root inputs. In humid regions, warming removes the temperature limitation on microbial respiration, accelerating decomposition and leading to net SOC losses, especially from the

more decomposed POM pool. Deeper-rooted native grass plants can help maintain or even increase soil carbon reserves in the face of climate change if they're managed using appropriate forms of grazing (Bai & Cotrufo, 2022).

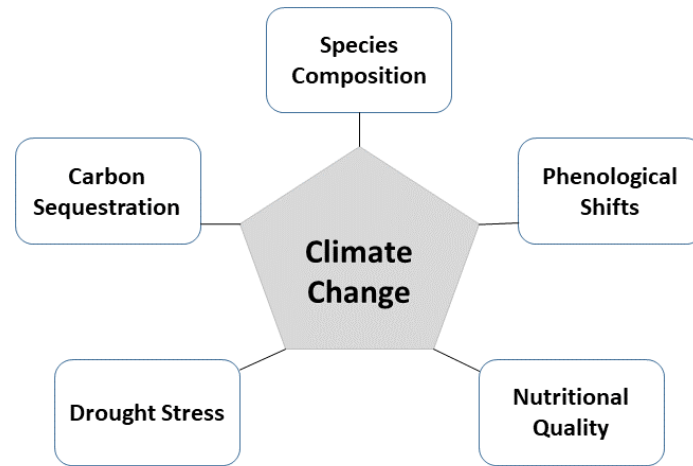


Figure 1.5. The various effects of climate change on grasslands in the Prairie Ecozone (PE).

1.3. The Landscape of Manitoba's PE

Three of the seven ecoregions in PE, namely, Aspen Parkland, Lake Manitoba Plain, and Southwest Manitoba Upland, are present in Manitoba (Figure 1.6). Aspen Parkland, a unique Canadian ecoregion, is found in the province's southwest, with a separate area located between the Riding and Duck Mountains. The mean annual air temperature in this ecoregion ranges from 1.2 °C to 2.8 °C. The average growing season days ranges from 173 days in the north to 183 in the south, with the number of growing degree days ranging from 1470 to 1700 days. The average annual precipitation ranges from 440 to 530 mm and exhibits significant yearly fluctuations. The growing season accounts for the largest proportion of precipitation. In Manitoba, annual moisture deficits range from under 140 mm in the north to around 300 mm in the southwest. The elevations change from 320 meters above sea level (masl) near the Manitoba Escarpment to around 600 masl in the hilly uplands to the west, which creates an eastward slope and is drained by the Souris, Assiniboine, Qu'Appelle and Pembina rivers. The different vegetation types found in the ecoregions include trembling aspen and shrubs on moist sites, and bur oak and grassland communities on increasingly drier sites. Dominant grasses in this ecoregion include fescues, wheat-grasses, June-grass and Kentucky bluegrass, accompanied by a diverse range of deciduous shrubs and herbs. Poorly drained areas support

slough grasses, marsh reed grass, sedges, cattails, and shrubby willows (Shorthouse, 2010; Smith et al., 1998).

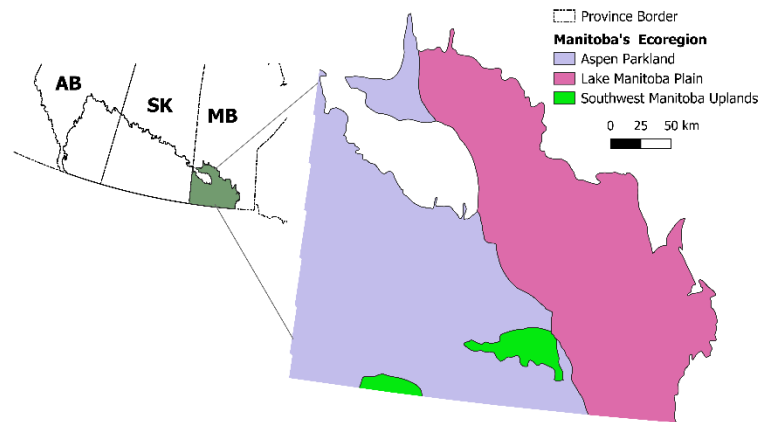


Figure 1.6. Prairie Ecozone (PE) in Manitoba and its ecoregions, including the Aspen Parkland, the Lake Manitoba Plain and the Southwest Manitoba Upland.

Within the Aspen Parkland Ecoregion, two areas are collectively recognized as a separate ecoregion, known as the Manitoba Upland Ecoregion. This ecoregion comprises two distinct uplands, the Turtle Mountain Upland, which forms the southern part and spans the border with the United States, and the Pembina Hills Upland, located within the Aspen Parkland, which forms the northern section of the ecoregion in south-central Manitoba (Smith et al., 1998). Before settlement, the hills were predominantly forested with trembling aspen and bur oak (Smith et al., 1998). Oak trees were distributed on the drier and warmer slopes, while extremely dry sites supported shrubs interspersed with grasses and herbs. The Pembina Hills, which rise approximately 200 meters above the adjacent Manitoba Plain to the east, form a warmer subdivision and serve as an outlier of the Boreal Ecoclimatic Region in Manitoba. The Turtle Mountain, which is associated with the Ecodistrict, serves as an outlier of the Boreal Plains Ecozone to the north and is a distinguished wooded upland. Agriculture is much more prevalent in the Pembina Hills area than in the Turtle Mountain area. The average annual precipitation varies from 540 mm in the Pembina Hills section to about 600 mm in the Turtle Mountain section (Shorthouse, 2010; Smith et al., 1998).

The most eastern ecoregion of the PE is the Lake Manitoba Plain, which includes Winnipeg. This ecoregion stretches northwestward from the US border to Dauphin Lake. It is recognized as one of the warmest and most moist regions in the Canadian prairies. The average annual air temperature varies between 2°C in the north and 3°C along the Canada-

United States border. The average temperature in summer is 16°C, while in winter it is -12.5°C. The annual precipitation varies between 450 and 700 mm. The elevation of the plain also varies from approximately 410 masl near the Manitoba Escarpment to 218 masl at Lake Winnipeg. The ecoregion areas were once naturally covered with aspen trees, fescue grasses, and various shrubs, but this vegetation disappeared due to cultivation. Tallgrass is one of the most endangered ecosystems in North America, reduced to 4 km²—about 0.07% of its original 6000 km²—in the Red River Basin (Shorthouse, 2010; Smith et al., 1998).

The dominant grass species in the remaining tallgrass prairies consist of big bluestem, switchgrass, Indian-grass, little bluestem, Canada wild rye, and northern wheat-grass. Big bluestem is a tall, C4, sod-forming grass known for its high nutritional value and adaptability to various conditions. Its colour changes from bluish-green in summer to reddish-brown in fall. This species typically grows between 1 and 2 m tall, with a fibrous root system extending up to 2 m deep. Switchgrass is most commonly found in prairie lowlands, while little bluestem is a medium-height, C4 bunchgrass common on prairie uplands (Shorthouse, 2010).

Besides grasslands in Manitoba's PE, other land cover such as wetlands, rivers, lakes, prairies, aspen parkland, shrubs, and broadleaf deciduous forest can be found. The main rivers in the region are the Assiniboine and Red Rivers. The Red River originates from the U.S. and flows northward, emptying into Lake Winnipeg. Its adjacent area was the location of the first agricultural colonies in the region (Hanuta, 2006). The Assiniboine River originates in eastern Saskatchewan province, flows through Manitoba's Prairie and eventually joins the Red River before emptying into Lake Winnipeg. Wetlands develop in areas with poor drainage; parklands form in areas that get more rain or where water flows from the surface; and deciduous forests are located along streams and have some coniferous trees at higher levels (Hanuta, 2006). The largest land cover class is associated with agricultural production, such as wheat, canola, soybeans, and potatoes. According to the Manitoba Grassland Inventory (MGI), cropland area was approximately 29,900 km² in the Prairie region in 2021 (Agriculture | Province of Manitoba, 2021).

Historically, grasslands only consisted of native grasses, but today they consist of native, tame and mixed classes. Ecologists have classified three distinct Prairie types in Manitoba, including tall grass Prairie, mixed grass Prairie and fescue Prairie (The Manitoba Museum, 2012). Tall grasses have mostly declined so far, and much of the remaining mixed

grass Prairie has been invaded by spreading aspen and, in some areas, bur oak (Scott, 1996). Fescue Prairie is dominated by fescue grasses and is found in areas that are moister than mixed grass Prairies. It is predominantly found in the Aspen Parkland, except in the southwestern part of Manitoba (The Manitoba Museum, 2012).

Chernozems are the main type of soil and have a thick surface layer full of organic matter (Smith et al., 1998). These soils have developed under native prairie vegetation, where a significant portion of plant biomass exists below ground. This results in the accumulation of organic matter throughout the soil profile, leading to the formation of dark, fertile topsoil layers (Manitoba Agriculture, Food and Rural Initiatives, 2008). In most cases, the reduction of organic matter and the thinning of the organic-rich layer in the soil have resulted from cultivation. In some instances, it is believed that the current organic content is only around 50 percent of what it was before (Smith et al., 1998). Nevertheless, the organic matter in these soils keeps them naturally fertile and ensures they hold much water. Hence, they are still very useful for farming. The Manitoba Prairie region consists of the best soils, including the Black, Dark Grey and Dark Brown Chernozems of the Aspen Parkland and the Lake Manitoba Plain (Smith et al., 1998). The water deficit during the year is the main reason for the differences in Chernozemic soil colour. Most of the region's dry spots have a water shortage of about 200 mm, causing native grasses to be low and have less biomass above and below ground. The small amount of organic matter coming from plants leads to soil that is grouped as Brown Chernozemic. Being farther west and north of the dry region reduces the lack of water, which in turn increases organic carbon and leads to the creation of Dark Brown and Black Chernozemic soils. The top levels of soil organic carbon exist in southern Manitoba, a region with an annual water deficit of about 70 mm (Pennock, 2021).

In terms of land morphology, as discussed earlier regarding the properties of each ecoregion, the elevation variation of Manitoba's PE ranges from 218 to 760 masl. As shown in Figure 1.7, the elevation map of PE can be divided geographically into two major areas: the eastern part with a minimum elevation of around 200 masl, and the western part with more height fluctuation, ranging from around 340 to 760 masl. In this part, higher elevation regions are distributed primarily around Turtle Mountain, the south and west of Riding Mountain and west of Virden.

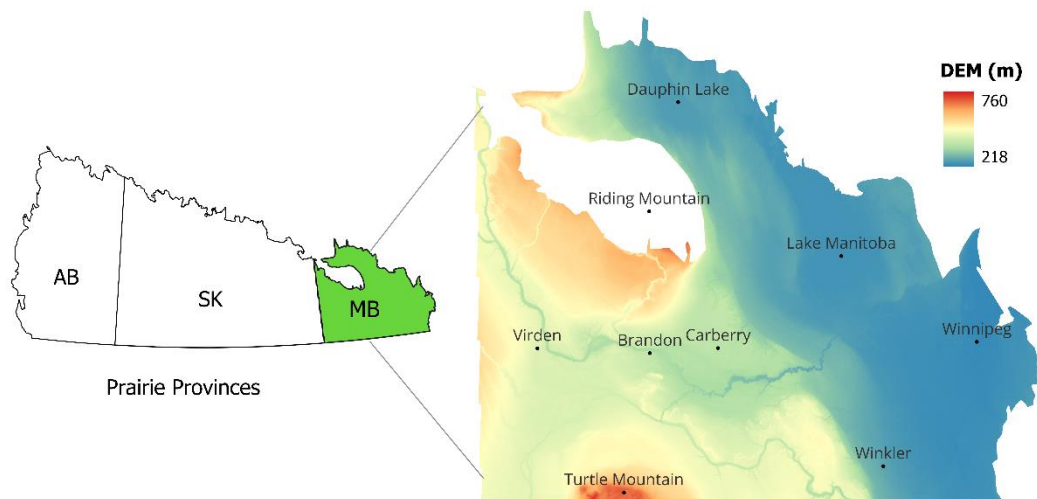


Figure 1.7. Digital elevation model (DEM) of Prairie Ecozone (PE) in the province of Manitoba, ranging from 218 to 760 meters. Elevation data were retrieved from SRTM with 30 m spatial resolution.

1.4. Grassland Mapping Using Remote Sensing Technology

Grassland classification is a preliminary step for further understanding these valuable ecosystems (Wu et al., 2023). The classification of grassland types forms the foundation for making decisions, reasonable evaluation, and the utilization of grassland resources (Sun et al., 2024). Timely and precise data about the grassland classification is indispensable in the sustainable planning and successful management of grassland ecosystems (Wu et al., 2023). Therefore, reliable grassland surveys, both temporally and spatially, can provide valuable insights into the quantity and quality characteristics of grasslands, which are essential for the rational development and conservation of these ecosystems (Sun & Ongsomwang, 2023). By accurately measuring the spatial distribution and composition of grassland types, it is feasible to obtain important vegetation parameters such as grassland yield and degradation status (Sun et al., 2024). This information supports the development of effective grassland utilization strategies. By timely classification of grasslands, it is possible to track temporal changes in grassland composition, such as the transition between C3 and C4 grasses or the spread of tame versus native species (Badreldin et al., 2021; Peterson et al., 2002). Agricultural expansion, human activities such as urban growth and overgrazing, and climate change are significant threats to the grassland biome. These changes can be detected using land cover classification, and their long-term trends can be monitored over time. Analyzing the direction and rate of these changes enables the prediction of future landscape dynamics based on the identified driving factors (Zandler et al., 2022; Zhang et al., 2023).

Grassland dynamics are closely linked to understanding changes in gross primary productivity (GPP) and net primary productivity (NPP) (Wang et al., 2022). NPP represents the difference between the total energy fixed by autotrophs through photosynthesis (GPP) and the energy they consume through respiration (R) (Z. Wang et al., 2022). Getting an accurate estimate of NPP is essential for evaluating the amount of vegetation growth and the potential for carbon sequestration in terrestrial ecosystems. (Lei et al., 2020). In native grassland ecosystems, NPP is related to the amount of AGB, giving information on how productive the area is and about the availability of forage and the rate of land degradation (Piipponen et al., 2022). That is why surveillance of NPP helps scholars learn about the ecosystem quality, carbon movement and effective use of grass pastures (Piipponen et al., 2022). NPP is directly tied to AGB accumulation and provides useful knowledge about ecosystem productivity, forage availability, and land degradation status (Zhao et al., 2014).

AGB is a reliable indicator of grassland health (Jia et al., 2024). Human activities have significant effects on the ability of grassland ecosystems to function, especially the quantity of plant biomass (Wei et al., 2024). These activities incorporate land use changes, disrupt nutrient cycles, and change species compositions. Such complex impacts make it difficult to assess their effects precisely (Liu et al., 2007). To have sustainable development and realize the terrestrial ecosystem dynamics, quantifying their effects on ecosystem functions is fundamental (Wei et al., 2024). Human activities include both positive interferences, such as managed grazing (Dong et al., 2022), and negative impacts, such as overgrazing and forest clearance (Bardgett et al., 2021), that can affect grassland ecosystem functions. AGB variations can show the direct influences of human activities on the grassland (Wei et al., 2024). Thus, monitoring AGB can help characterize the spatial and temporal effects of human activities on the grassland ecosystem.

Both grassland classification and quantifying grassland AGB offer reliable and beneficial insights for various applications in agriculture, ecology, environmental monitoring and economic planning, all of which are fundamental for sustainable improvements. Grassland, as one of the most extensive biomes, faces challenges for traditional approaches that often fail to track the temporal changes and spatial variation across large areas. Time and cost are major limitations when analyzing such a large ecosystem over different periods. Remote

sensing offers an efficient solution to overcome these constraints by enabling wide-scale and timely observation of the Earth's surface.

Several types of sensors have been used to monitor Earth phenomena so far, including terrestrial and aquatic ecosystems. Optical remote sensing is a major group of sensors that are widely used for various applications, particularly in grassland monitoring (Wang et al., 2022). This category can be divided further into subgroups like multispectral, high-spatial resolution and hyperspectral remote sensing (Zhu et al., 2017). Moderate Resolution Imaging Spectrometer (MODIS), Landsat and Sentinel are some examples of multispectral sensors. Their spatial resolution ranges from 250 to 10 meters, meaning that each pixel corresponds to a square area on the ground with the same size as the sensor's nominal resolution (Zhu et al., 2017). Satellite data are essential for identifying various grassland types and observing changes in vegetation and biomass over time. Remote sensing technology provides an efficient, cost-effective way to map and monitor these vast ecosystems, overcoming the limitations of traditional field-based methods, which are often time-consuming and labour-intensive (Wang et al., 2022). Additionally, remote sensing can provide real-time data, helping track dynamic changes in grasslands caused by factors like climate change and human activities, such as overgrazing (Zeng et al., 2021).

Grasslands and their dynamics have been evaluated using various sensors and different platforms (e.g., satellites, airplanes and drones), but multispectral sensors are more commonly used due to their low cost and open accessibility. The first MODIS instrument was launched with the Terra satellite in 1999, and the second sensor operated in 2002 with the Aqua satellite. They captured images at 36 spectral bands at varying spatial resolutions (two bands at 250 m, five bands at 500 m and 29 bands at 1 km). MODIS data, with 1-2 days temporal resolution, have been typically utilized for regional and global applications in grassland monitoring (Gao et al., 2013; Jia et al., 2016). Landsat and Sentinel are popular multispectral sensors that are freely available for users. Landsat has been monitoring the Earth for many years, as each new version replaces the one before it. Currently, Landsat 8 and Landsat 9 are the last operational sensors in the Landsat series (launched in 2013 and 2021, respectively) (Markham et al., 2020). These sensors operate similarly, with a 16-day revisit time, and a spatial resolution of 30 meters for Visible, near infrared (NIR), and shortwave infrared (SWIR) bands. Meanwhile, the optical Sentinel sensor, which has been working since

2015, offers improved spatial resolution (10 m for Visible and NIR, and 20 m for SWIR) and a shorter repeated observation period of 5 days (Phiri et al., 2020). Wang et al. (2022) discovered that studies making use of MODIS data for grassland observations are greater in number than those employing data from Landsat and Sentinel sensors. This situation is linked to a variety of factors. MODIS has gathered images of Earth for longer than Landsat 8, Landsat 9 and the optical Sentinel missions, which have been operating for around a decade. Because grassland ecosystems are distributed across a wide range of areas, it is practical to use a remote sensing sensor that covers a large area, and MODIS excels at this. MODIS offers multiple valuable vegetation products, including the NDVI product MOD13Q1, which is directly applied in numerous studies (Wang et al., 2022).

However, for national and local scales, both Landsat and Sentinel are more practical, as they provide higher spatial resolution and greater detail (Hargreaves & Watmough, 2021). With the recent and rapid development of cloud-based platforms such as Google Earth Engine (GEE), Amazon Web Services (AWS), SEPAL, and NASA NEX, the use of Landsat and Sentinel data for even larger-scale applications has become more feasible. Managing and processing large volumes of data has been significantly simplified using these cloud-based tools (Hargreaves & Watmough, 2021; Parracciani et al., 2024).

Hyperspectral imagery is a special type of multispectral remote sensing that measures continuous spectrum, resulting in hundreds of narrow spectral bands (Amigo, 2019). These images have often been utilized for grassland classifications (Dao et al., 2021), AGB estimation (Shen et al., 2008), management (Askari et al., 2019; Lyu et al., 2020) and production (Geipel & Korsath, 2017). High (meters) and ultra-high (centimetres) spatial resolution images provide a novel data source for exploring environmental questions with exceptional detail (He & Weng, 2018). These high-resolution data are the results of significant progress in platforms and sensors. With higher spatial resolution, intraclass variability has increased, making it more challenging to analyze and classify images. For example, in grassland imagery with high resolution, the details of leaves, stems, gaps and shadows can all be seen in pixels of one scene. While this extra detail may improve how accurate maps are, it also adds difficulties for segmenting images and choosing features (He & Weng, 2018). Lu and He (2018) utilized ultra-high spatial resolution of unmanned aerial vehicle (UAV) imagery to investigate the species

composition in a heterogeneous grassland ecosystem. High-resolution data were also used in grassland AGB modelling (Wang et al., 2014) and grassland use intensities (Asam et al., 2015).

Besides optical remote sensing, another major type is radar or microwave remote sensing, which are usually active sensors imaging in the electromagnetic wavelength range from 1 mm to 1 m (Moreira et al., 2013). Microwave signals are emitted to the Earth's surface and, after scattering with targets, backscattered to the antenna of the sensor and record the physical and dielectric behaviour (Moreira et al., 2013). The radar imaging is side-looking, with the platform moving along the flight path while illuminating its footprint obliquely. The main characteristic of radar sensors is that they are weather-independent; whether it is rainy or foggy weather, they can capture images. They can also penetrate targets, including soils, water, forest and so on (Zhu et al., 2017). Radar imaging systems transmit a microwave signal that is either horizontally polarized (H) or vertically polarized (V). Polarization refers to the orientation of the electric field. Some radar sensors, such as PALSAR, are capable of transmitting and receiving microwave signals at both polarization modes (H and V), making them fully polarimetric sensors.

1.5. The Development in Spaceborne Technology

Observing Earth from space, which provides a broader view, started in the past few decades, with the dawn of satellite technology. Since then, numerous remote sensing approaches have been developed (Wang et al., 2022). These methods are beneficial for applications that require monitoring extensive areas and repetitive data to track changes in dynamic themes, such as agriculture (Palanisamy et al., 2019), water quality (Gholizadeh et al., 2016), and sea temperature (Merchant, 2013). A brief overview of remote sensing principles is provided to support the methodological approach.

Remote sensing is any technique used to gather information about an object that does not require touch or physical contact with the object. Remote sensing is composed of three elements, including the platform, target or object and the instrument or sensor (Campbell & Wynne, 2011). The primary requirement for operating a remote sensing system is to have a source of energy to illuminate the objects. This energy, which is in the form of electromagnetic radiation, can be produced by sensors or by the sun as a source of energy or radiation (Campbell & Wynne, 2011). Thus, based on the source of energy remote sensing

sensors can be divided into two groups: (i) active sensors when sensors emit their electromagnetic radiation and measure the reflected or scattered signal [examples include RADAR (Radio Detection and Ranging), LiDAR (Light Detection and Ranging)], and (ii) passive sensors when sensors detect natural electromagnetic radiation emitted or reflected by objects, typically from the sun [examples include optical, multispectral, hyperspectral, and thermal infrared sensors] (Canada Centre for Remote Sensing, 2010). Two examples of optical and radar imagery, which were acquired by Sentinel-2 and Sentinel-1 sensors, respectively, around Winnipeg are shown in Figure 1.8.

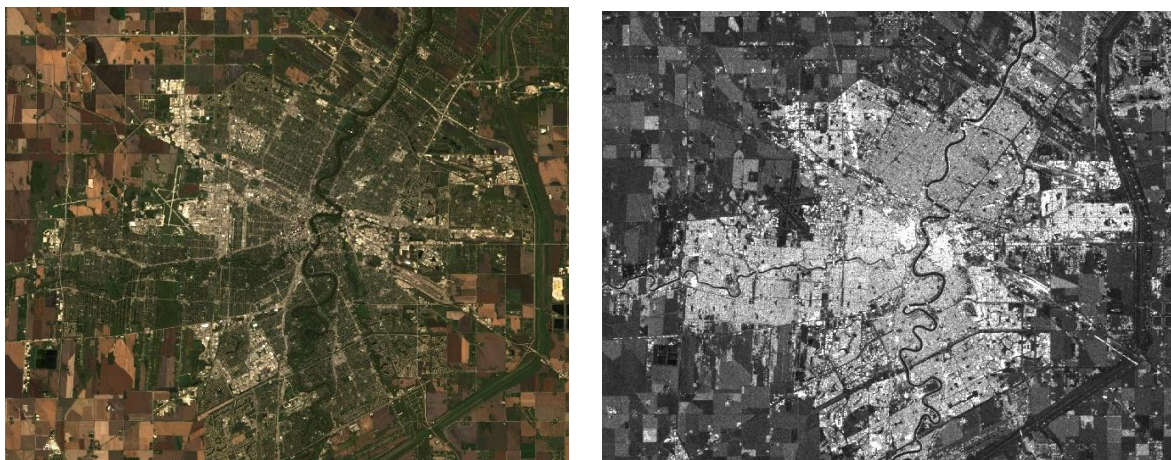


Figure 1.8. Remote sensing imagery of Winnipeg and neighbouring areas using (left) the Passive Sentinel-2 sensor and (right) the Active Sentinel-1 sensor.

Orbits or paths of satellites that carry sensor(s) are carefully designed to match the sensor's objective and capability. Satellites that observe a portion of Earth continuously have geostationary orbits. Weather monitoring and communications satellites follow these types of orbits. Many remote sensing platforms follow the north-south orbits to cover the majority of the Earth's surface. These orbits are called near-polar orbits, also known as sun-synchronous orbits (Campbell & Wynne, 2011). It is because these satellites in these orbits pass over the same location on Earth at approximately the same local solar time each day. This means that the angle of sunlight remains consistent for every pass, ensuring uniform lighting conditions for imaging. In these orbits, when the satellite moves northwards, it is on its ascending pass, and then moving toward the southern pole is known as the descending pass (Campbell & Wynne, 2011).

The era of repetitive and systematic Earth observations began with the launch of the first multispectral scanner system (Landsat 1) in 1972. To date, many multispectral sensors,

e.g., the Landsat series, MODIS, High Resolution Visible and Infrared (HRVIR), the SPOT family, IKONOS, GeoEye sensors, and the Sentinel constellation, have been utilized to derive data for various applications, such as environmental assessment and global change detection, mapping, military surveillance, and ecosystem monitoring (Badreldin et al., 2017; Wang et al., 2020). However, remote sensing methods face various challenges involving spatial and temporal resolution, spectral bandwidth, and data deluge (Wang et al., 2022).

1.5.1. Satellite Image Characteristics

The distance between objects on Earth and the spaceborne platforms determines how much detail the sensor can capture and the total area covered in each image. Spatial resolution has a direct relationship with the height of the sensor. When a sensor has a spatial resolution of 10 meters, meaning each pixel in the image represents a 10 by 10-meter square on the ground. The required resolution is determined based on the scale of the application, whether local, neighbourhood, national, or continental. Images with highly detailed features are considered to have fine or high resolution; those with low detail are referred to as coarse or low resolution. In general, the finer resolution images capture more details but cover a smaller area on the ground (Chuvieco, 2020).

Spectral resolution describes how precisely a sensor can detect and distinguish different wavelengths of light. Different types of features and details can be differentiated by comparing their spectral responses across specific wavelength bands. For example, to easily distinguish between different rock types, a sensor with higher spectral resolution is required. In contrast to multispectral sensors that have several bands, hyperspectral sensors capture hundreds of narrow bands (Chuvieco, 2020).

The radiometric properties define the potential of a sensor to distinguish different levels of grey scale (Figure 1.9). When an image is captured by a sensor, the sensor's sensitivity to electromagnetic energy influences its radiometric resolution. The radiometric resolution indicates how precisely a sensor can measure the intensity of the electromagnetic radiation it receives (Rani et al., 2023). As images are recorded in digital format, the level of grey scale values is measured in bits and bytes. For instance, a sensor with 8-bit radiometric resolution can record 256 levels of intensity ($2^8 = 256$), or a 12-bit sensor can detect 4,096 levels ($2^{12} = 4,096$). Generally, finer radiometric resolution results in larger data volumes (Rani et al., 2023).

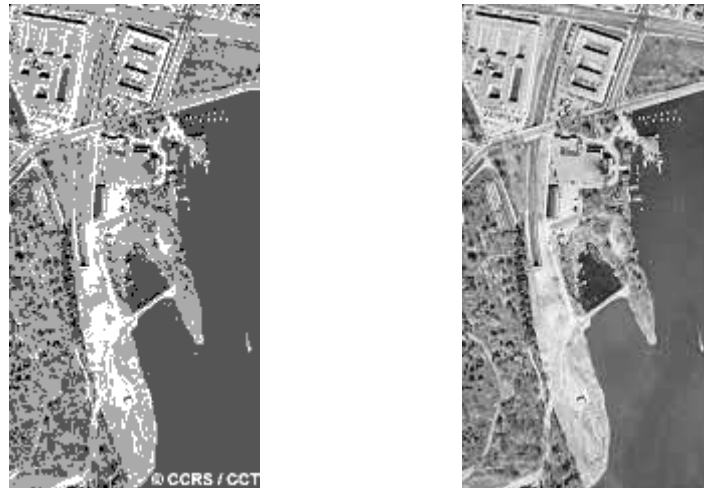


Figure 1.9. Image contrasting lower (left) and higher (right) radiometric resolution, showing the greater level of grey scale (Canada Centre for Remote Sensing, 2010.)

Temporal resolution refers to how often a sensor revisits and captures data from the same area on Earth's surface. This frequency is determined by the sensor's characteristics and the satellite's orbital pattern. Depending on the sensor's role, temporal resolution can vary. For example, Landsat 8 has a revisit time of approximately 16 days, while other sensors like MODIS can provide near-daily observations. The temporal resolution plays a pivotal role in monitoring changes taking place in a specific area on Earth over time. These changes, whether natural or human-induced, often result in different spectral characteristics that can facilitate the detection of certain features. For example, it enables the measurement of urban development over time or helps distinguish between similar crops, such as wheat and maize, based on their different growth patterns (Kotawadekar, 2021).

1.6. The Application of Artificial Intelligence in Monitoring Grasslands

Artificial intelligence (AI) is rapidly advancing the field of grassland monitoring, offering new tools for assessing grassland health, productivity, and ecological changes (Diouf et al., 2016; Iqbal, 2024; Morais et al., 2021; Pinna et al., 2024). Integrating AI with remote sensing and digital infrastructure (e.g., high-performance computers and quantum computers) allows researchers and land managers to monitor grassland ecosystems more accurately and frequently. AI includes technologies enabling machines to mimic human cognitive abilities, while machine learning, a subset of AI, allows systems to automatically improve through experience (Google, 2025).

Machine learning algorithms enhance the accuracy and efficiency of grassland classification and AGB estimation when applied to data obtained from remote sensing. Machine learning models, such as random forest (RF), support vector machine (SVM), and artificial neural networks (ANN), can process complex, high-dimensional datasets and uncover patterns that traditional methods might miss (Dusseux et al., 2022; Morais et al., 2021). These algorithms are particularly useful for analyzing large datasets that cover extensive areas. Machine learning models can be trained using labelled data (i.e., a training set) to recognize specific grassland types and estimate aboveground biomass (AGB) values, facilitating continuous and accurate monitoring of grassland ecosystems. As more high-quality remote sensing data is available and machine learning can process non-linear relationships, using both technologies together greatly helps in improving grassland management, protecting them and monitoring the carbon cycle (Fan et al., 2022; Mohajane et al., 2021). ANN, especially back-propagation models, is used to estimate AGB with high accuracy. These models utilize field measurements and remote sensing data, providing detailed spatial coverage and outperforming traditional regression methods ($R^2 = 0.75-0.85$ versus $0.40-0.64$) (Zeng et al., 2021).

As humans learn from experience, machines can learn from experience and improve their performance through computational methods. In computer systems, data or observations play the role of experience (Zhou, 2021). By developing learning algorithms and creating models, the model is able to make predictions about an occurrence (e.g., biomass, cropland classification, and CO₂ emission). Depending on the type of prediction output, whether it is discrete or continuous, it is called a classification or regression problem (Zhou, 2021). Machine learning methods can be broadly divided into two main classes: supervised learning and unsupervised learning (Zocca et al., 2017).

In supervised learning, the learning process is achieved by previously labelled data, so supervised learning algorithms can predict similar but unlabelled data (Zhou, 2021). This type of learning algorithm may be used for classification and/or regression problems (Lary et al., 2016), based on whether the output is discrete or continuous values [Equation (1)].

$$f: X (\text{observations}) \rightarrow y (\text{classes} = \text{discrete values or continuous values}) \quad (1)$$

Training data are the collection of observations or features used to teach machine learning models how to make predictions. Depending on whether they are labelled or not, they are used for supervised or unsupervised problems, respectively. Training data should cover a wide range of scenarios (i.e., system parameters space) to ensure the model learns diverse patterns and relationships (Lary et al., 2016). The process of analyzing the prediction model using new data is testing. A random subset of the data is used as testing data to check the model functionality (also called validation) (Lary et al., 2016).

Error rate is the proportion of misclassified samples (f) to the total number of samples (s). Accordingly, accuracy is $1 - \frac{f}{s}$ or $1 - \text{error rate}$. In general terms, the difference between predicted values generated by machine learning models and corresponding ground truth or true values is called error and indicates model performance. While this error is calculated using training data, it is called training error or empirical error, and while the error is calculated using test data, it is called generalization error. Two key problems that affect machine learning performance are overfitting and underfitting. Overfitting occurs when the learner model learns the training data “too well”, making it perform excellently on training data but poorly on new data. On the opposite, when the learner model is too simple to capture underlying patterns in training data (Zhou, 2021).

There are several types of machine learning models, like RF, SVM and ANN, that were used in this research and in the following sub-sections, an overview of each model will be provided.

1.6.1. Decision Trees (DTs)

DTs are one of the popular machine learning methods used for classification and regression. Each DT consists of decision nodes and leaf nodes (Figure 1.10).

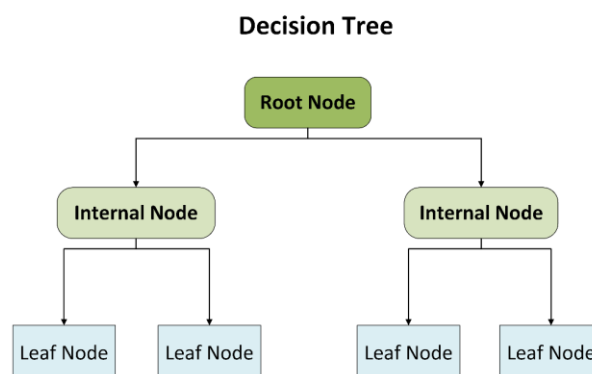


Figure 1.10. A complete decision tree (DT) model, where there is a hierarchical tree structure that begins with one main question at the top, called the root node. Internal decisions are new decision nodes that were leaf nodes before.

Selecting the optimal splitting feature is the key step in the DT learning algorithm. Ideally, as splitting progresses, the purity of each node should increase, meaning that most samples in a node belong to a single class. The Gini index and entropy are two main approaches used to measure the node purity in DTs (Zhou, 2021). The Gini index, as shown in Equation (2), is computationally more efficient as it measures node impurity. A lower Gini value indicates higher node purity, meaning the samples predominantly belong to a single class (Hastie et al., 2009).

$$Gini(D) = 1 - \sum_{i=1}^n (P_i)^2 \quad (2)$$

where, P_i is the probability of an object being classified into a particular class.

DTs suffer from several key limitations, including overfitting, instability and class imbalance issues, which impact their performance and reliability (Géron, 2022). Ensemble methods that utilize DTs as base learners (e.g., RFs, Gradient Boosted approaches) effectively address the key limitations of individual DTs (Géron, 2022). Gradient tree boosting (GTB) and RF are both popular ensemble learning techniques used in machine learning for classification and regression tasks. The current ensemble methods can be divided into two main types based on how the base models are generated (Zhou, 2021). The first type, illustrated by Bagging and RF, produces independent learners that can be generated in parallel. The second type, exemplified by Boosting, develops highly interdependent learners sequentially (Zhou, 2021). RF constructs multiple DTs independently. Each tree is built on a randomly selected subset of the training data (bootstrapping) and a random subset of features. The final prediction is determined by a majority vote among all trees (Géron, 2022). GTB builds DTs sequentially. Each new tree in the ensemble focuses on reducing the errors made by the previous trees. The algorithm fits each new tree on the residual errors of the previous ensemble (Géron, 2022). Both algorithms are illustrated in Figure 1.11 for better understanding.

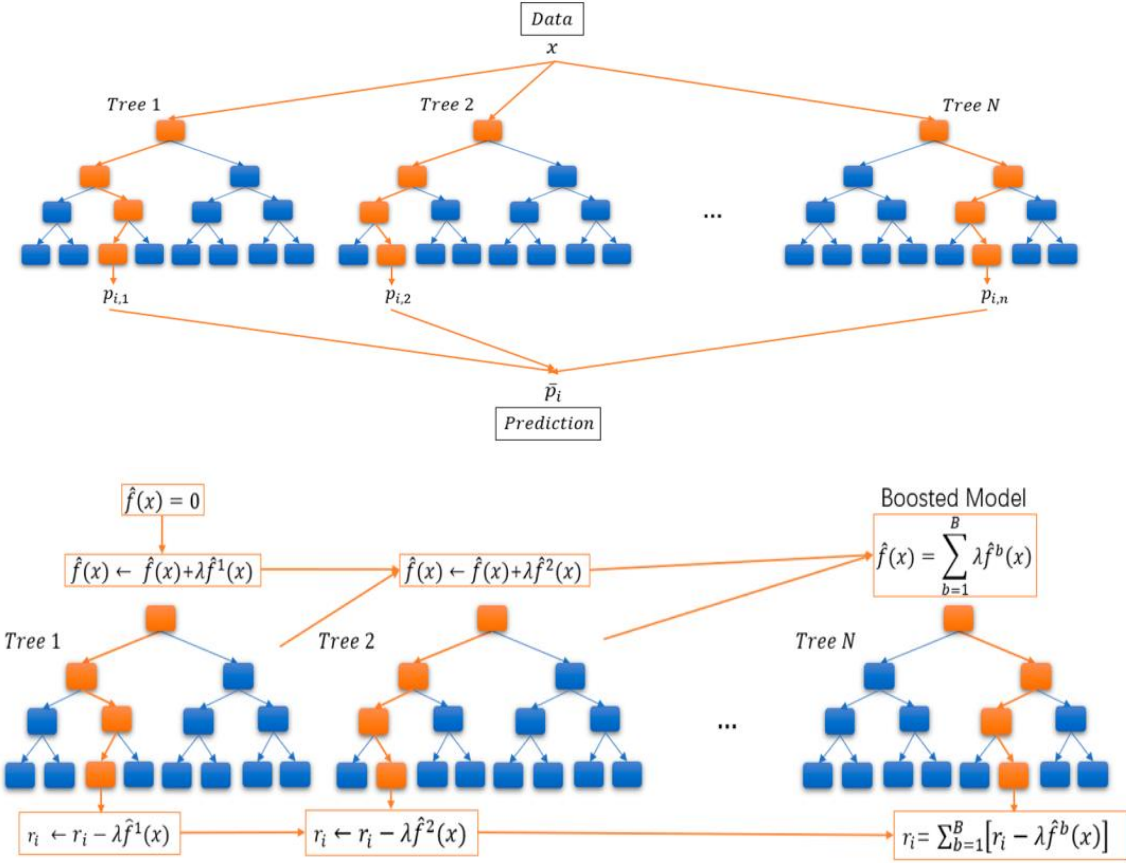


Figure 1.11. Graphical illustration of random forest (RF; top) and gradient tree boosting (GTB; bottom) algorithms (Jiao, Hu, et al., 2020; Jiao, Sun, et al., 2020).

1.6.2. SVM

SVMs are a supervised learning method used for classification and regression tasks (Zocca et al., 2017). They work by finding the optimal hyperplane that maximizes the margin between different classes in the feature space (Zocca et al., 2017). The equation of a hyperplane can be expressed as the following linear function [Equation (3)]:

$$W^T X + b = 0 \quad (3)$$

where $W=(w_1, w_2, \dots, w_d)$ is the weight vector that is normal to the hyperplane, and b is the bias. The coefficients (w, b) control the hyperplane (Zhou, 2021). In Figure 1.12, a 2D graph is displayed in which the hyperplane is a straight line. As shown, the best hyperplane is the plane that has the maximum distance from two groups of points (Zhou, 2021). The distance from any point to the hyperplane (w, b) can be expressed by Equation (4):

$$d = \frac{|w^b x + b|}{\|w\|} \quad (4)$$

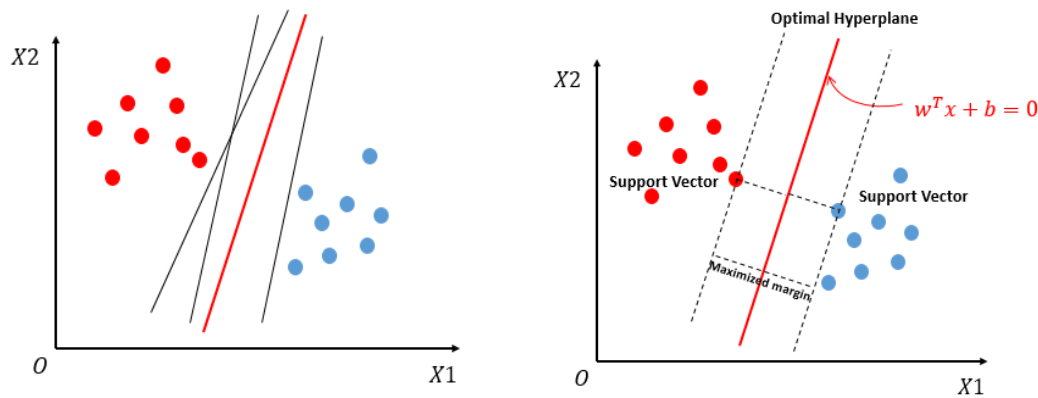


Figure 1.12. Illustration of support vector machine algorithm. Left: there are several hyperplanes to separate blue balls from red balls, but only one optimal hyperplane with maximum margin. Right: optimal support vectors and margins.

If sample data are not linearly separable, SVM can deal with such cases through two key ways: First, using soft margins. It works by intentionally allowing some misclassified training examples to improve the model's generalization ability (Zocca et al., 2017). By comforting the strict linear reparability requirement, this approach prevents overfitting and enhances robustness to noise. The second approach is using the kernel trick that maps the input space into a higher-dimensional feature space where linear separation becomes possible, using kernel functions as illustrated in Figure 1.13 (Zocca et al., 2017).

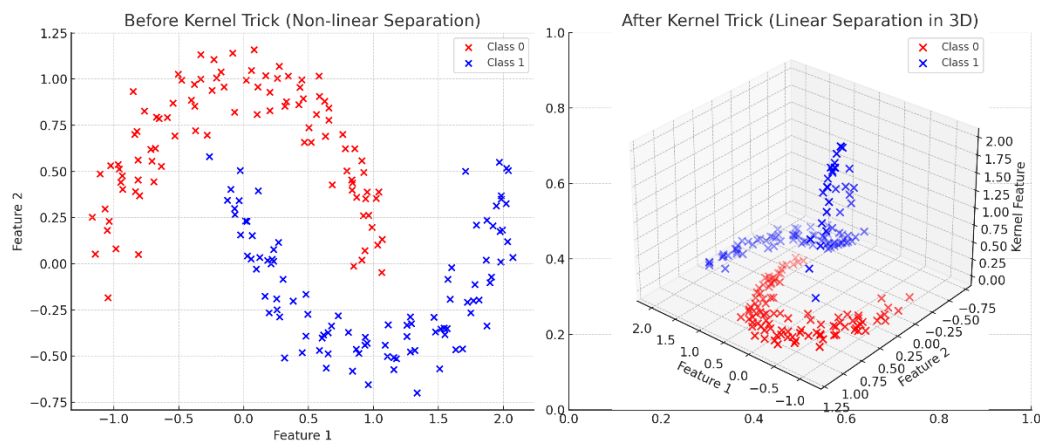


Figure 1.13. Illustration of kernel separation in support vector machines (SVMs). In the left graph, a non-linearly separable dataset was shown before the kernel was applied. On the right, the same dataset after the kernel has been applied, and the classes become linearly separable using a hyperplane.

1.6.3. ANN

ANN was first formulated in 1943 by McCulloch and Pitts (Zhou, 2021). They were inspired by the biological function of neurons in the human brain (Poddar, 2024). Over the decades, advancements in technology, particularly in information technology (IT), have

significantly impacted artificial intelligence (Russell & Norvig, 2016). These developments, as well as the emergence of large-scale datasets, have driven the evolution of neural networks, from early single-layer perceptrons (Rosenblatt, 1958) to modern architectures like convolutional neural networks (CNNs) (Kelleher, 2019). The key components of artificial neural networks comprise neurons, layers, weights, biases, activation functions, and training processes (Poddar, 2024). The basic units are neurons that receive inputs and are governed by a threshold and an activation function (Figure 1.14).

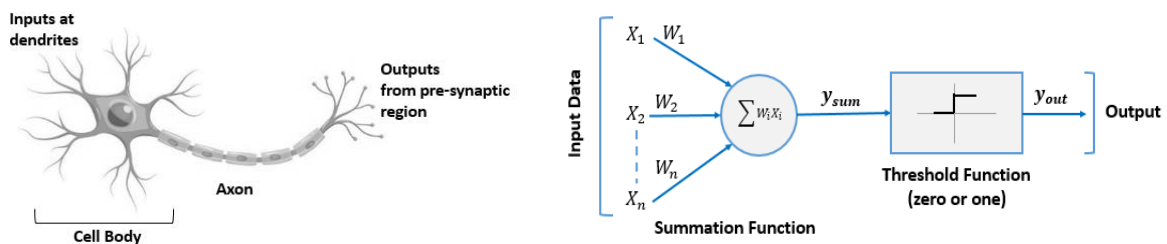


Figure 1.14. Illustration of an artificial neural network. The left figure shows the input and output structure of a biological neuron (Poddar, 2024). On the right, McCulloch-Pitts (or perceptron) provides a decision/output based on input data (Graupe, 2013). Its activation function and weighted inputs mimic the functions of biological axons and dendrites, respectively (Graupe, 2013).

ANNs consist of layers, including input, output, and hidden layers, with greater layer depth indicating higher model complexity (Figure 1.15). Each connection between neurons carries information, regulated by weights and biases. These weights and biases determine the influence of one neuron on another. During the training process, the values of biases and weights are adjusted. Activation functions play a vital role by introducing non-linearity into the system, enabling neural networks to capture sophisticated patterns in data. Popular activation functions include Rectified Linear Unit (ReLU), Sigmoid, and Tanh. The training process involves regulating the weights and biases to minimize the difference between the model's predictions and actual outputs. Typically, backpropagation integrated with optimization techniques like gradient descent is used to update weights and biases (Poddar, 2024).

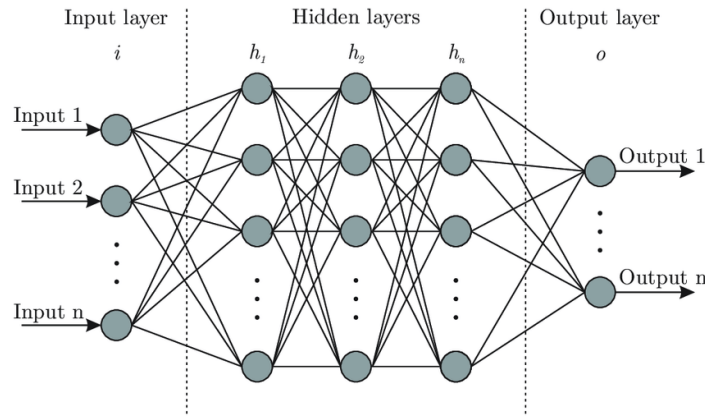


Figure 1.15. The architecture of an artificial neural network (ANN) model (Vawda et al., 2024).

1.7. Thesis Objectives

Mapping grasslands within the Prairie region of Manitoba was a major goal of this research (Chapter 2). A novel hierarchical AI-based strategy was designed to differentiate between three distinct grassland classes (native, mixed, and tame grasses). Apart from two classifications conducted by government organizations, specifically annual crop inventory (ACI) and MGI, there is a lack of comparative research on mapping grasslands in Manitoba's PE. A specific section of this study aims to fill this gap by providing a comparative classification map for the region.

There were also several sub-goals within the classification map, including handling different remote sensing sensors, such as Sentinel-1, Sentinel-2, Landsat 8-9, and evaluating their capabilities in land cover classification. Finding an optimal data combination and using multi-temporal imagery were key components of satellite data processing aimed at improving grassland classification accuracy. Meanwhile, the study assessed different classification machine learning methods, including RF, SVM and GTB, in terms of grassland classification performance. Finally, two fusion techniques, such as hue, saturation, and value (HSV) and Brovey transformation (BT), were evaluated, as it was not possible to use multi-temporal imagery, and their effects on grassland classification accuracy were examined.

The second major goal of this research was to estimate grassland AGB using remote sensing data (Chapter 3). While several studies worldwide have focused on remote sensing-based methods to estimate AGB in grasslands, this study was the first to specifically utilize remote sensing data and machine learning techniques to assess grassland AGB in Manitoba's Prairie Ecozone (PE). This study aimed to assess various relationships between spectral indices

and in-situ biomass samples. Several sub-objectives supported the main objective of Chapter 3. First, three optical remote sensing sensors, including PlanetScope, Sentinel-2 and Landsat 8-9, featuring various spatial and temporal resolutions, were used to evaluate their effectiveness in estimating grassland AGB. Meanwhile, several types of statistical models, including parametric models [i.e., univariate and multivariate linear and exponential regression models] and machine learning methods [i.e., RF, SVM, and ANN], were assessed for estimating biomass. Additionally, a regression model was used to populate biomass samples, facilitating the generation of synthetic samples throughout the entire growing season. Finally, the effect of environmental features like precipitation, temperature and elevation on grassland estimation was examined.

1.8. Structure of Thesis

This thesis includes two manuscript-format research chapters, prepared following the thesis guidelines set out by the Department of Soil Science at the University of Manitoba. Chapter 1 provides a general introduction to the concept of grasslands and their ecosystem services in Canada, particularly in Manitoba, as well as an overview of machine learning approaches, remote sensing technology, and the objectives of the research. Chapter 2 focuses on the approaches used to classify grasslands in Manitoba's PE. A grasslands map is generated to show the spatial distribution of native, tame and mixed grasses. Chapter 3 explains how grassland AGB is estimated using remote sensing data. In this chapter, parametric and machine learning models are assessed to estimate grassland AGB using various remote sensing data with different spatial resolutions. In addition, auxiliary data like climatic factors and elevation variable are also evaluated in biomass estimation. Chapter 4 highlights the research findings and offers recommendations for future research development using novel tools and datasets.

1.9. References

- Agriculture | Province of Manitoba. (2021). Manitoba grassland inventory (MGI) – land identification project. <https://www.gov.mb.ca/agriculture/>
- Agriculture | Province of Manitoba. (2024). Forage seed production in manitoba. <https://www.gov.mb.ca/agriculture/>
- Akinremi, O., McGinn, S., & Cutforth, H. (2001). Seasonal and spatial patterns of rainfall trends on the Canadian prairies. *Journal of Climate*, 14(9), 2177–2182.
- Amigo, J. M. (2019). *Hyperspectral imaging* (Vol. 32). Elsevier.
- Asam, S., Klein, D., & Dech, S. (2015). Estimation of grassland use intensities based on high spatial resolution LAI time series. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7-W3, 285–291. 36th International Symposium on Remote Sensing of Environment (Volume XL-7/W3) - 11–15 May 2015, Berlin, Germany. <https://doi.org/10.5194/isprsarchives-XL-7-W3-285-2015>
- Askari, M. S., McCarthy, T., Magee, A., & Murphy, D. J. (2019). Evaluation of grass quality under different soil management scenarios using remote sensing techniques. *Remote Sensing*, 11(15), 1835.
- Badreldin, N., Prieto, B., & Fisher, R. (2021). Mapping Grasslands in Mixed Grassland Ecoregion of Saskatchewan Using Big Remote Sensing Data and Machine Learning. *Remote Sensing*, 13(24), Article 24. <https://doi.org/10.3390/rs13244972>
- Badreldin, N., Xing, Z., & Goossens, R. (2017). The application of satellite-based model and bi-stable ecosystem balance concept to monitor desertification in arid lands, a case study of Sinai Peninsula. *Modeling Earth Systems and Environment*, 3(1), 21. <https://doi.org/10.1007/s40808-017-0300-5>
- Bai, Y., & Cotrufo, M. F. (2022). Grassland soil carbon sequestration: Current understanding, challenges, and solutions. *Science*, 377(6606), 603–608. <https://doi.org/10.1126/science.abo2380>
- Bardgett, R. D., Bullock, J. M., Lavorel, S., Manning, P., Schaffner, U., Ostle, N., Chomel, M., Durigan, G., L. Fry, E., Johnson, D., Lavallee, J. M., Le Provost, G., Luo, S., Png, K., Sankaran, M., Hou, X., Zhou, H., Ma, L., Ren, W., ... Shi, H. (2021). Combatting global grassland degradation. *Nature Reviews Earth & Environment*, 2(10), 720–735. <https://doi.org/10.1038/s43017-021-00207-2>
- Bazzaz, F. A., & Parrish, J. A. D. (1982). Organization of grassland communities. In *Grasses and grassland communities: Systematics and ecology* (Estes, J. R. Tyrl, R. J. Brunken, J. N., pp. 233–254). University of Oklahoma Press.
- Belesky, D. P., & Fedders, J. M. (1995). Comparative Growth Analysis of Cool- and Warm-Season Grasses in a Cool–Temperate Environment. *Agronomy Journal*, 87(5), 974–980. <https://doi.org/10.2134/agronj1995.00021962008700050034xdup>
- Bengtsson, J., Bullock, J. M., Egoh, B., Everson, C., Everson, T., O'Connor, T., O'Farrell, P. J., Smith, H. G., & Lindborg, R. (2019). Grasslands—More important for ecosystem

- services than you might think. *Ecosphere*, 10(2), Article 2. <https://doi.org/10.1002/ecs2.2582>
- Blair, J., Nippert, J., & Briggs, J. (2014). Grassland ecology 14. *Ecology and the Environment*, 389, 389–423.
- Bonsal, B. R., Zhang, X., Vincent, L. A., & Hogg, W. D. (2001). *Characteristics of Daily and Extreme Temperatures over Canada*. https://journals.ametsoc.org/view/journals/clim/14/9/1520-0442_2001_014_1959_codaet_2.0.co_2.xml
- Bush, E., Lemmen, D. S., & editors. (2019). Canada's Changing Climate Report (p. 444). Government of Canada, Ottawa, ON. <https://changingclimate.ca/CCCR2019/>
- Cai, T., Chang, C., Zhao, Y., Wang, X., Yang, J., Dou, P., Otgonbayar, M., Zhang, G., Zeng, Y., & Wang, J. (2024). Within-season estimates of 10 m aboveground biomass based on Landsat, Sentinel-2 and PlanetScope data. *Scientific Data*, 11(1), 1276. <https://doi.org/10.1038/s41597-024-04120-3>
- Campbell, J. B., & Wynne, R. H. (2011). Introduction to remote sensing. Guilford press.
- Canada Centre for Remote Sensing. (2010). *Fundamentals of Remote Sensing*. Natural Resources Canada.
- Cao, J., Zhao, H., & Qin, Y. (2019). Biomass allocation between above- and below-ground and it impacted factors of shrubby areas and grasslands in upper heihe river basin of China. *AIP Conference Proceedings*, 2079(1), Article 1. <https://doi.org/10.1063/1.5092404>
- Carrier, L., & Bort, K. S. (1916). The History of Kentucky Bluegrass and White Clover in the United States. *Agronomy Journal*, 8(4), 256–267. <https://doi.org/10.2134/agronj1916.00021962000800040007x>
- Chakravarti, A. (1972). The June-July precipitation pattern in the prairie provinces of Canada. *Journal of Geography*, 71(3), 155–160.
- Cook, I. (2019, October 18). Grassland Birds at Risk. *Birds Canada | Oiseaux Canada*. <https://www.birdscanada.org/bird-science/grassland-birds-at-risk>
- Chuvieco, E. (2020). Fundamentals of satellite remote sensing: An environmental approach. CRC press.
- Climate Atlas of Canada. (2022). <https://climateatlas.ca/home-page>
- Craine, J. M., Nippert, J. B., Elmore, A. J., Skibbe, A. M., Hutchinson, S. L., & Brunsell, N. A. (2012). Timing of climate variability and grassland productivity. *Proceedings of the National Academy of Sciences*, 109(9), 3401–3405. <https://doi.org/10.1073/pnas.1118438109>
- Crews, T. E., & Rumsey, B. E. (2017). What agriculture can learn from native ecosystems in building soil organic matter: A review. *Sustainability*, 9(4), 578.
- Cutforth, H., & Judiesch, D. (2012). Temperature trends in the semiarid prairie of southwestern Saskatchewan revisited. *Canadian Journal of Soil Science*, 92(5), 803–806. <https://doi.org/10.4141/cjss2012-015>

- Cutforth, H. W., & Judiesch, D. (2007). Long-term changes to incoming solar energy on the Canadian Prairie. *Agricultural and Forest Meteorology*, 145(3), 167–175. <https://doi.org/10.1016/j.agrformet.2007.04.011>
- Dao, P. D., Axiotis, A., & He, Y. (2021). Mapping native and invasive grassland species and characterizing topography-driven species dynamics using high spatial resolution hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102542. <https://doi.org/10.1016/j.jag.2021.102542>
- Dass, P., Houlton, B. Z., Wang, Y., & Warlind, D. (2018). Grasslands may be more reliable carbon sinks than forests in California. *Environmental Research Letters*, 13(7), 074027. <https://doi.org/10.1088/1748-9326/aacb39>
- Diouf, A. A., Hiernaux, P., Brandt, M., Faye, G., Djaby, B., Diop, M. B., Ndione, J. A., & Tychon, B. (2016). Do agrometeorological data improve optical satellite-based estimations of the herbaceous yield in Sahelian semi-arid ecosystems? *Remote Sensing*, 8(8), 668.
- Dong, L., Zheng, Y., Martinsen, V., Liang, C., & Mulder, J. (2022). Effect of Grazing Exclusion and Rotational Grazing on Soil Aggregate Stability in Typical Grasslands in Inner Mongolia, China. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.844151>
- Dusseux, P., Guyet, T., Pattier, P., Barbier, V., & Nicolas, H. (2022). Monitoring of grassland productivity using Sentinel-2 remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 111, 102843. <https://doi.org/10.1016/j.jag.2022.102843>
- ECCC. (2025). Climate Trends and Variations Bulletin. <https://www.canada.ca/en/environment-climate-change/services/climate-change/science-research-data/climate-trends-variability/trends-variations.html>
- Fan, X., He, G., Zhang, W., Long, T., Zhang, X., Wang, G., Sun, G., Zhou, H., Shang, Z., Tian, D., Li, X., & Song, X. (2022). Sentinel-2 Images Based Modeling of Grassland Above-Ground Biomass Using Random Forest Algorithm: A Case Study on the Tibetan Plateau. *Remote Sensing*, 14(21), Article 21. <https://doi.org/10.3390/rs14215321>
- Forage Species. (2024). *Beef Research*. <https://www.beefresearch.ca/topics/forage-species/>
- Gao, T., Xu, Bin, Yang, Xiuchun, Jin, Yunxiang, Ma, Hailong, Li, Jinya, & Yu, H. (2013). Using MODIS time series data to estimate aboveground biomass and its spatio-temporal variation in Inner Mongolia's grassland between 2001 and 2011. *International Journal of Remote Sensing*, 34(21), 7796–7810. <https://doi.org/10.1080/01431161.2013.823000>
- Geipel, J., & Korsæth, A. (2017). Hyperspectral Aerial Imaging for Grassland Yield Estimation. *Advances in Animal Biosciences*, 8(2), 770–775. <https://doi.org/10.1017/S2040470017000619>
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc.

- Ghahramani, A., Howden, S. M., del Prado, A., Thomas, D. T., Moore, A. D., Ji, B., & Ates, S. (2019). Climate Change Impact, Adaptation, and Mitigation in Temperate Grazing Systems: A Review. *Sustainability*, 11(24), Article 24. <https://doi.org/10.3390/su11247224>
- Gholizadeh, M. H., Melesse, A. M., & Reddi, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors*, 16(8), Article 8. <https://doi.org/10.3390/s16081298>
- Google. (2025). AI vs. Machine Learning: How Do They Differ? Google Cloud. <https://cloud.google.com/learn/artificial-intelligence-vs-machine-learning>
- Government of Canada, P. S. and P. C. (2002, July 1). *Prairies Ecozone+ evidence for key findings summary: Canadian biodiversity - ecosystem status and trends 2010. Evidence for key findings summary report No. 4 / published by the Canadian Councils of Resource Ministers. : En14-43/0-4-2014E-PDF - Government of Canada Publications - Canada.ca*. <https://publications.gc.ca/site/eng/9.629546/publication.html>
- Graupe, D. (2013). *Principles Of Artificial Neural Networks (3rd Edition)*. World Scientific.
- Hanuta, I. (2006). *Land cover and climate for part of southern Manitoba: A reconstruction from Dominion Land Survey maps and historical records of the 1870s*.
- Hargreaves, P. K., & Watmough, G. R. (2021). Satellite Earth observation to support sustainable rural development. *International Journal of Applied Earth Observation and Geoinformation*, 103, 102466. <https://doi.org/10.1016/j.jag.2021.102466>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Havrilla, C. A., Bradford, J. B., Yackulic, C. B., & Munson, S. M. (2023). Divergent climate impacts on C3 versus C4 grasses imply widespread 21st century shifts in grassland functional composition. *Diversity and Distributions*, 29(3), 379–394. <https://doi.org/10.1111/ddi.13669>
- He, Y., & Weng, Q. (2018). *High spatial resolution remote sensing: Data, analysis, and applications*. CRC press.
- Iqbal, M. A. (2024). Introductory Chapter: Present and Future of Artificial Intelligence in Grasslands Conservation. In *Grasslands—Conservation and Development*. IntechOpen. <https://doi.org/10.5772/intechopen.114190>
- Jia, W., Liu, M., Yang, Y., He, H., Zhu, X., Yang, F., Yin, C., & Xiang, W. (2016). Estimation and uncertainty analyses of grassland biomass in Northern China: Comparison of multiple remote sensing data sources and modeling approaches. *Ecological Indicators*, 60, 1031–1040. <https://doi.org/10.1016/j.ecolind.2015.09.001>
- Jia, Z., Zhang, Z., Cheng, Y., Buhebaoyin, Borjigin, S., & Quan, Z. (2024). Grassland biomass spatiotemporal patterns and response to climate change in eastern Inner Mongolia based on XGBoost model estimates. *Ecological Indicators*, 158, 111554. <https://doi.org/10.1016/j.ecolind.2024.111554>

- Jiao, Z., Hu, P., Xu, H., & Wang, Q. (2020). Machine Learning and Deep Learning in Chemical Health and Safety: A Systematic Review of Techniques and Applications. *ACS Chemical Health & Safety*, 27(6), 316–334. <https://doi.org/10.1021/acs.chas.0c00075>
- Jiao, Z., Sun, Y., Hong, Y., Parker, T., Hu, P., Mannan, M. S., & Wang, Q. (2020). Development of Flammable Dispersion Quantitative Property–Consequence Relationship Models Using Extreme Gradient Boosting. *Industrial & Engineering Chemistry Research*, 59(33), 15109–15118. <https://doi.org/10.1021/acs.iecr.0c02822>
- Kelleher, J. D. (2019). *Deep learning*. MIT press.
- Knapp, A. K., Briggs, J. M., Hartnett, D. C., & Collins, S. L. (Eds.). (1998). *Grassland Dynamics: Long-Term Ecological Research in Tallgrass Prairie* (1st edition). Oxford University Press.
- Kotawadekar, R. (2021). 9 - Satellite data: Big data extraction and analysis. In D. Binu & B. R. Rajakumar (Eds.), *Artificial Intelligence in Data Mining* (pp. 177–197). Academic Press. <https://doi.org/10.1016/B978-0-12-820601-0.00008-2>
- Lark, T. J., Spawn, S. A., Bougie, M., & Gibbs, H. K. (2020). Cropland expansion in the United States produces marginal yields at high costs to wildlife. *Nature Communications*, 11(1), 4295. <https://doi.org/10.1038/s41467-020-18045-z>
- Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, 7(1), 3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>
- Lee, M. A., Davis, A. P., Chagunda, M. G. G., & Manning, P. (2017). Forage quality declines with rising temperatures, with implications for livestock production and methane emissions. *Biogeosciences*, 14(6), 1403–1417. <https://doi.org/10.5194/bg-14-1403-2017>
- Lei, T., Feng, J., Lv, J., Wang, J., Song, H., Song, W., & Gao, X. (2020). Net Primary Productivity Loss under different drought levels in different grassland ecosystems. *Journal of Environmental Management*, 274, 111144. <https://doi.org/10.1016/j.jenvman.2020.111144>
- Liu, H., Mi, Z., Lin, L., Wang, Y., Zhang, Z., Zhang, F., Wang, H., Liu, L., Zhu, B., Cao, G., Zhao, X., Sanders, N. J., Classen, A. T., Reich, P. B., & He, J.-S. (2018). Shifting plant species composition in response to climate change stabilizes grassland primary production. *Proceedings of the National Academy of Sciences*, 115(16), 4051–4056. <https://doi.org/10.1073/pnas.1700299114>
- Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H., & Taylor, W. W. (2007). Complexity of Coupled Human and Natural Systems. *Science*, 317(5844), 1513–1516. <https://doi.org/10.1126/science.1144004>
- Lu, B., & He, Y. (2018). Optimal spatial resolution of Unmanned Aerial Vehicle (UAV)-acquired imagery for species classification in a heterogeneous grassland ecosystem. *GIScience & Remote Sensing*, 55(2), 205–220. <https://doi.org/10.1080/15481603.2017.1408930>

- Lyu, X., Li, X., Dang, D., Dou, H., Xuan, X., Liu, S., Li, M., & Gong, J. (2020). A new method for grassland degradation monitoring by vegetation species composition using hyperspectral remote sensing. *Ecological Indicators*, 114, 106310. <https://doi.org/10.1016/j.ecolind.2020.106310>
- Manitoba Agriculture, Food and Rural Initiatives. (2008). *Soil Management Guide*.
- Mapfumo, E., Chanasyk, D. S., Puurveen, D., Elton, S., & Acharya, S. (2023). Historic climate change trends and impacts on crop yields in key agricultural areas of the prairie provinces in Canada: A literature review. *Canadian Journal of Plant Science*, 103(3), 243–258. <https://doi.org/10.1139/cjps-2022-0215>
- Markham, B., Jenstrom, D., Sauer, B., Pszcolka, S., Dulski, V., Hair, J., McCorkel, J., Kvaran, G., Thome, K., Montanaro, M., Pedelty, J., Anderson, C., Choate, M., Barsi, J., Kaita, E., & Miller, J. (2020). Landsat 9 mission update and status. *Earth Observing Systems XXV*, 11501, 96–102. <https://doi.org/10.1117/12.2569748>
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, Ö., Yu, R., & Zhou, B. (Eds.). (2021). IPCC. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 3–32). Cambridge University Press. <https://doi.org/10.1017/9781009157896.001>
- McGinn, S. M. (2010). Weather and climate patterns in Canada's prairie grasslands. *Arthropods of Canadian Grasslands*, 1, 105–119.
- McGinn, S., & Shepherd, A. (2003). Impact of climate change scenarios on the agroclimate of the Canadian prairies. *Canadian Journal of Soil Science*, 83(5), 623–630.
- Merchant, C. J. (2013). Thermal Remote Sensing of Sea Surface Temperature. In C. Kuenzer & S. Dech (Eds.), *Thermal Infrared Remote Sensing: Sensors, Methods, Applications* (pp. 287–313). Springer Netherlands. https://doi.org/10.1007/978-94-007-6639-6_15
- Milner, C. (1968). *Methods for the measurement of the primary production of grassland*. Blackwell Scientific.
- Mohajane, M., Costache, R., Karimi, F., Bao Pham, Q., Essahlaoui, A., Nguyen, H., Laneve, G., & Oudija, F. (2021). Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area. *Ecological Indicators*, 129, 107869. <https://doi.org/10.1016/j.ecolind.2021.107869>
- Morais, T. G., Teixeira, R. F. M., Figueiredo, M., & Domingos, T. (2021). The use of machine learning methods to estimate aboveground biomass of grasslands: A review. *Ecological Indicators*, 130, 108081. <https://doi.org/10.1016/j.ecolind.2021.108081>
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1), 6–43. <https://doi.org/10.1109/MGRS.2013.2248301>
- Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F. A., Schwarz, L.-M., Ghazaryan, G., Schultz, J., Malenovský, Z., & Dubovyk, O. (2022). Predicting plant biomass and

- species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. *Remote Sensing of Environment*, 282, 113262. <https://doi.org/10.1016/j.rse.2022.113262>
- National Grassland Inventory. (2024). Canadian Forage & Grassland Association. <https://www.canadianfga.ca/en/projects/grassland-inventory/>
- Palanisamy, S., Selvaraj, R., Ramesh, T., & Ponnusamy, J. (2019). Applications of Remote Sensing in Agriculture—A Review. *International Journal of Current Microbiology and Applied Sciences*, 8, 2270–2283. <https://doi.org/10.20546/ijcmas.2019.801.238>
- Parracciani, C., Gigante, D., Bonini, F., Grassi, A., Morbidini, L., Pauselli, M., Valenti, B., Lilli, E., Antonielli, F., & Vizzari, M. (2024). Leveraging Google Earth Engine for a More Effective Grassland Management: A Decision Support Application Perspective. *Sensors*, 24(3), Article 3. <https://doi.org/10.3390/s24030834>
- Pennock, D. J. (2021). *Soils of the Prairie Provinces*. <https://www.saskoer.ca/soilscience/chapter/soils-of-the-prairie-provinces/>
- Petermann, J. S., & Buzhdygan, O. Y. (2021). Grassland biodiversity. *Current Biology*, 31(19), R1195–R1201. <https://doi.org/10.1016/j.cub.2021.06.060>
- Peterson, D. L., Price, K. P., & Martinko, E. A. (2002). Discriminating between cool season and warm season grassland cover types in northeastern Kansas. *International Journal of Remote Sensing*, 23(23), 5015–5030. <https://doi.org/10.1080/01431160210142833>
- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing*, 12(14), Article 14. <https://doi.org/10.3390/rs12142291>
- Piipponen, J., Jalava, M., de Leeuw, J., Rizayeva, A., Godde, C., Cramer, G., Herrero, M., & Kummu, M. (2022). Global trends in grassland carrying capacity and relative stocking density of livestock. *Global Change Biology*, 28(12), 3902–3919. <https://doi.org/10.1111/gcb.16174>
- Pinna, D., Pezzuolo, A., Cogato, A., Pornaro, C., Macolino, S., & Marinello, F. (2024). Applications of satellite platforms and machine learning for mapping and monitoring grasslands and pastures: A systematic and comprehensive review. *Smart Agricultural Technology*, 9, 100571. <https://doi.org/10.1016/j.atech.2024.100571>
- Poddar, H. (2024). From neurons to networks: Unravelling the secrets of artificial neural networks and perceptrons. In *Deep Learning in Engineering, Energy and Finance* (pp. 25–79). CRC Press.
- Pogue, S. J., Kröbel, R., Janzen, H. H., Beauchemin, K. A., Legesse, G., de Souza, D. M., Iravani, M., Selin, C., Byrne, J., & McAllister, T. A. (2018). Beef production and ecosystem services in Canada’s prairie provinces: A review. *Agricultural Systems*, 166, 152–172. <https://doi.org/10.1016/j.agsy.2018.06.011>
- Prairie Grasslands. (2014, September 5). *Celebrating Manitoba’s Provincial Grass Emblem*. <https://manitobagrass.ca/about-prairie-grasslands/>

- Rani, A., Kumari, A., Pinky, & Kumar, M. (2023). Chapter 13—Aspects of forest degradation and inventory approaches for forest management. In U. Chatterjee, B. Pradhan, S. Kumar, S. Saha, M. Zakwan, B. D. Fath, & D. Fiscus (Eds.), *Water, Land, and Forest Susceptibility and Sustainability* (Vol. 2, pp. 363–387). Academic Press. <https://doi.org/10.1016/B978-0-443-15847-6.00015-X>
- Ren, H., Eviner, V. T., Gui, W., Wilson, G. W. T., Cobb, A. B., Yang, G., Zhang, Y., Hu, S., & Bai, Y. (2018). Livestock grazing regulates ecosystem multifunctionality in semi-arid grassland. *Functional Ecology*, 32(12), 2790–2800. <https://doi.org/10.1111/1365-2435.13215>
- Roch, L., & Jaeger, J. A. G. (2014). Monitoring an ecosystem at risk: What is the degree of grassland fragmentation in the Canadian Prairies? *Environmental Monitoring and Assessment*, 186(4), 2505–2534. <https://doi.org/10.1007/s10661-013-3557-9>
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. pearson.
- Schüle, M., Heinken, T., & Fartmann, T. (2023). Long-term effects of environmental alterations in protected grasslands – Land-use history determines changes in plant species composition. *Ecological Engineering*, 188, 106878. <https://doi.org/10.1016/j.ecoleng.2022.106878>
- Scott, G. A. (1996). Manitoba's ecoclimatic regions. *The Geography of Manitoba, Its Land and Its People*, 43–59.
- Shen, M., Tang, Y., Klein, J., Zhang, P., Gu, S., Shimono, A., & Chen, J. (2008). Estimation of aboveground biomass using in situ hyperspectral measurements in five major grassland ecosystems on the Tibetan Plateau. *Journal of Plant Ecology*, 1(4), 247–257. <https://doi.org/10.1093/jpe/rtn025>
- Shorthouse, J. D. (2010). Ecoregions of Canada's prairie grasslands. *Arthropods of Canadian Grasslands*, 1, 53–81.
- Skinner, W., & Gullett, D. (1993). Trends of daily maximum and minimum temperature in Canada during the past century. *Climatological Bulletin*, 27(2), 63–77.
- Smith, R. E., Veldhuis, H., Mills, G. F., Eilers, R. G., Fraser, W. R., & Lelyk, G. W. (1998). *Terrestrial Ecozones, Ecoregions and Ecodistricts of Manitoba*.
- Srivastava, R. K., & Yetgin, A. (2024). An overall review on influence of root architecture on soil carbon sequestration potential. *Theoretical and Experimental Plant Physiology*, 36(2), 165–178. <https://doi.org/10.1007/s40626-024-00323-6>
- Sun, B., Qin, P., Li, C., Gao, Z., Grainger, A., Li, X., Wang, Y., & Yue, W. (2024). Integrating vegetation phenological characteristics and polarization features with object-oriented techniques for grassland type identification. *Geo-Spatial Information Science*, 27(3), 794–810. <https://doi.org/10.1080/10095020.2023.2250378>
- Sun, J., & Ongsomwang, S. (2023). Optimal parameters of random forest for land cover classification with suitable data type and dataset on Google Earth Engine. *Frontiers in Earth Science*, 11. <https://doi.org/10.3389/feart.2023.1188093>

- The Manitoba Museum, T. M. (2012, November 12). Prairie Pollination. *Prairie Pollination / Prairie Pollinisation*. https://prairiepollination.ca/bienvenue_dans_les_prairies-welcome_to_the_prairies/types_de_prairie-prairie_types/
- Tilman, D., & El Haddi, A. (1992). Drought and biodiversity in Grasslands. *Oecologia*, 89(2), 257–264. <https://doi.org/10.1007/BF00317226>
- Tilman, D., Reich, P. B., & Knops, J. M. H. (2006). Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature*, 441(7093), 629–632. <https://doi.org/10.1038/nature04742>
- Touzi, R., & Shimada, M. (2009). Polarimetric PALSAR Calibration. *IEEE Transactions on Geoscience and Remote Sensing*, 47(12), 3951–3959. <https://doi.org/10.1109/TGRS.2009.2032176>
- Vawda, M. I., Lottering, R., Mutanga, O., Peerbhay, K., & Sibanda, M. (2024). Comparing the Utility of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) on Sentinel-2 MSI to Estimate Dry Season Aboveground Grass Biomass. *Sustainability*, 16(3), Article 3. <https://doi.org/10.3390/su16031051>
- Vendramini, J. M. B., Silveira, M. L., & Moriel, P. (2023). Resilience of warm-season (C4) perennial grasses under challenging environmental and management conditions. *Animal Frontiers: The Review Magazine of Animal Agriculture*, 13(5), 16–22. <https://doi.org/10.1093/af/vfad038>
- Wang, H., Wang, C., Price, K. P., van der Merwe, D., & An, N. (2014). Modeling above-ground biomass in tallgrass prairie using ultra-high spatial resolution sUAS imagery. *Photogrammetric Engineering & Remote Sensing*, 80(12), 1151–1159.
- Wang, Y., Lu, Z., Sheng, Y., & Zhou, Y. (2020). Remote Sensing Applications in Monitoring of Protected Areas. *Remote Sensing*, 12(9), Article 9. <https://doi.org/10.3390/rs12091370>
- Wang, Z., Ma, Y., Zhang, Y., & Shang, J. (2022). Review of Remote Sensing Applications in Grassland Monitoring. *Remote Sensing*, 14(12), Article 12. <https://doi.org/10.3390/rs14122903>
- Wangchuk, K., Darabant, A., Nirola, H., Wangdi, J., & Gratzner, G. (2021). Climate Warming Decreases Plant Diversity but Increases Community Biomass in High-Altitude Grasslands. *Rangeland Ecology & Management*, 75, 51–57. <https://doi.org/10.1016/j.rama.2020.11.008>
- Watersheds. (2023, April 2). *Watersheds Canada*. <https://watersheds.ca/how-native-plants-help-with-erosion-control/>
- Wei, J.-Z., Pei, J.-Y., Fang, C., Sardans, J., Peñuelas, J., Huang, F.-Q., Zhang, X.-L., & Zheng, K. (2024). Detecting and attributing the impact of human activities on grassland aboveground biomass in China's Loess Plateau. *Journal of Environmental Management*, 371, 123247. <https://doi.org/10.1016/j.jenvman.2024.123247>
- Wheaton, E. E. (1998). *But it's a dry cold: Weathering the Canadian Prairies*. Calgary: Fifth House Publishers.

- Whitlaw, H. (2022, July 4). Grasslands Canada. *Sierra Club Canada*.
<https://www.sierraclub.ca/grasslands-canada/>
- Wu, N., Crusiol, L. G. T., Liu, G., Wuyun, D., & Han, G. (2023). Comparing Machine Learning Algorithms for Pixel/Object-Based Classifications of Semi-Arid Grassland in Northern China Using Multisource Medium Resolution Imageries. *Remote Sensing*, 15(3), Article 3. <https://doi.org/10.3390/rs15030750>
- Wu, Y., Du, Y., Liu, X., Wan, X., Yin, B., Hao, Y., & Wang, Y. (2023). Grassland biodiversity response to livestock grazing, productivity, and climate varies across biome components and diversity measurements. *Science of The Total Environment*, 878, 162994. <https://doi.org/10.1016/j.scitotenv.2023.162994>
- Zandler, H., Faryabi, S. P., & Ostrowski, S. (2022). Contributions to Satellite-Based Land Cover Classification, Vegetation Quantification and Grassland Monitoring in Central Asian Highlands Using Sentinel-2 and MODIS Data. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.684589>
- Zapisocki, Z., Murillo, R. de A., & Wagner, V. (2022). Non-Native Plant Invasions in Prairie Grasslands of Alberta, Canada. *Rangeland Ecology & Management*, 83, 20–30. <https://doi.org/10.1016/j.rama.2022.02.011>
- Zeng, N., Ren, X., He, H., Zhang, L., Li, P., & Niu, Z. (2021). Estimating the grassland aboveground biomass in the Three-River Headwater Region of China using machine learning and Bayesian model averaging. *Environmental Research Letters*, 16(11), 114020. <https://doi.org/10.1088/1748-9326/ac2e85>
- Zhang, X., Zhao, T., Xu, H., Liu, W., Wang, J., Chen, X., & Liu, L. (2023). GLC_FCS30D: The first global 30-m land-cover dynamic monitoring product with a fine classification system from 1985 to 2022 using dense time-series Landsat imagery and continuous change-detection method. *Earth System Science Data Discussions*, 2023, 1–32.
- Zhao, F., Xu, B., Yang, X., Jin, Y., Li, J., Xia, L., Chen, S., & Ma, H. (2014). Remote Sensing Estimates of Grassland Aboveground Biomass Based on MODIS Net Primary Productivity (NPP): A Case Study in the Xilingol Grassland of Northern China. *Remote Sensing*, 6(6), Article 6. <https://doi.org/10.3390/rs6065368>
- Zhou, Z.-H. (2021). *Machine Learning*. Springer Nature.
- Zhu, L., Suomalainen, J., Liu, J., Hyypä, J., Kaartinen, H., Haggren, H., Zhu, L., Suomalainen, J., Liu, J., Hyypä, J., Kaartinen, H., & Haggren, H. (2017). A Review: Remote Sensing Sensors. In *Multi-purposeful Application of Geospatial Data*. IntechOpen. <https://doi.org/10.5772/intechopen.71049>
- Zocca, V., Spacagna, G., Slater, D., & Roelants, P. (2017). *Python Deep Learning*. Packt Publishing Ltd.

2. MAPPING GRASSLAND IN MANITOBA'S PRAIRIE ECOZONE¹

2.1. Abstract

Accurate and reliable knowledge about grassland distribution is essential for farmers, stakeholders, and the government to manage grassland resources from agro-economic and ecological perspectives effectively. This study developed a novel pixel-based grassland classification approach using three supervised machine learning algorithms, which were assessed in the province of Manitoba, Canada. The grassland classification process involved three stages: (1) to distinguish between vegetation and non-vegetation covers, (2) to differentiate grassland from non-grassland landscapes, and (3) to identify three specific grassland classes (tame, native, and mixed grasses). Initially, this study investigated different satellite data, such as Sentinel-1, Sentinel-2, and Landsat 8 and 9, individually and combined, using the random forest (RF) method, with the best performance at the first two steps achieved using a combination of Sentinel-1 and Sentinel-2. The combination was then utilized to conduct the first two steps of classification using support vector machine (SVM) and gradient tree boosting (GTB). In step 3, after filtering out non-grassland pixels, the performance of RF, SVM, and GTB classifiers was evaluated with combined Sentinel-1 and Sentinel-2 data to distinguish different grassland types. Eighty-nine multitemporal raster-based variables, synthetic aperture radar (SAR) backscatters, and digital elevation model (DEM) were input for machine learning models. RF had the highest classification accuracy at 69.96% overall accuracy (OA) and a Kappa value of 0.55. After feature selection, the variables were reduced to 61, increasing OA to 72.62% with a Kappa value of 0.58. GTB ranked second, with its OA and Kappa values improving from 67.69% and 0.50 to 72.18% and 0.58 after feature selection. The impact of raster data quality on grassland classification accuracy was assessed through multisensor image fusion. Grassland classification using the Hue, Saturation, and Value (HSV) fused images showed higher OA (59.18%) and Kappa values (0.36) than the Brovey Transform (BT) and non-fused images. Finally, a web map was created to show grassland results within the Soil Landscapes of Canada (SLC) polygons, relating soil landscapes to grassland distribution and providing valuable information for decision-makers

¹ This chapter is based on the following published article: Mousavi, M., Biney, J. K. M., Kishchuk, B., Youssef, A., Cordeiro, M. R., Friesen, G., ... & Badreldin, N. (2024). A hierarchical machine learning-based strategy for mapping grassland in Manitoba's diverse ecoregions. *Remote Sensing*, 16(24), 4730. <https://doi.org/10.3390/rs16244730>

and researchers. Future work may include extending the current methodology by considering other influential variables, like meteorological parameters or soil properties, to create a comprehensive grassland inventory across the whole Prairie Ecozone of Canada.

2.2. Introduction

Grasslands account for around 40% of the Earth's terrestrial surface (Y. Wu et al., 2023) and are one of the most biologically diverse ecosystems (Roch & Jaeger, 2014). Globally, grasslands make a major contribution to the feed supply for livestock (Gibson, 2009). Soil and water conservation, carbon sequestration, and landscape biodiversity are other ecosystem services that grasslands provide (McInnes et al., 2015). The Prairie ecozone (PE) of Canada is an important and vulnerable ecozone due to its susceptibility to climate change and intensive agricultural production (Henderson & Sauchyn, 2008; Shorthouse, 2010). Native grasslands are degrading due to various stressors such as land use conversion to agriculture, grazing or recreation, the spread of invasive species, and climate change (Dao et al., 2021; Xu & Guo, 2015).

Mapping the distribution of native grasses is important to reverse degradation and promote biodiversity maintenance and species conservation (Godfree et al., 2017). Native grasses are important habitats for many wildlife species, including birds, mammals and insects. The loss of a significant amount of habitat has effects on the species that depend on it. The major threats to native grasslands in the Prairie regions include conversion of native grasslands to agricultural lands and the spread of non-native grasses due to anthropogenic interventions for economic benefits, resulting in the disappearance of local grasses over the past years (Mardian et al., 2021). For instance, Kentucky bluegrass (or blue grass) has been spreading across the Prairie that originally is native to Europe and North Asia (Carrier & Bort, 1916). Thus, monitoring the grassland ecosystem allows us to detect changes within the ecosystem and transitions between ecosystems when it interacts with the surrounding ecological systems.

Bi-temporal change detection allows for the identification of land use and land cover (LULC) changes in an area at different times. Ecosystems, particularly the grassland biome, are dynamic and human activities and climate changes can intensify the destructive or instructive alterations and damage the ecosystem's balance. This is particularly important for

ecologists, farmers, and stakeholders to track the trend in LULC changes. As shown in Figure 2.1, there were changes among some main LULC classes in Manitoba's Prairie region from 2012 to 2015. For example, pastures/forage areas have declined 60% (mainly changed to Grassland/Pasture class) while cropland has expanded 8%. The large change may be due to classification errors caused by spectrally similar classes, highlighting the importance of using a reliable classification method. This diagram was created using the Agriculture and Agri-Food Canada (AAFC) annual crop inventory (ACI) classification map and indicates the flow of changes over time. Land cover changes are generally detected using two broad approaches: pre-classification (which focuses on identifying changes in pixel values directly) and post-classification (which identifies changes in the classifications) (Basu & Das, 2021). In Figure 2.1, post-classification method was used with ACI data.

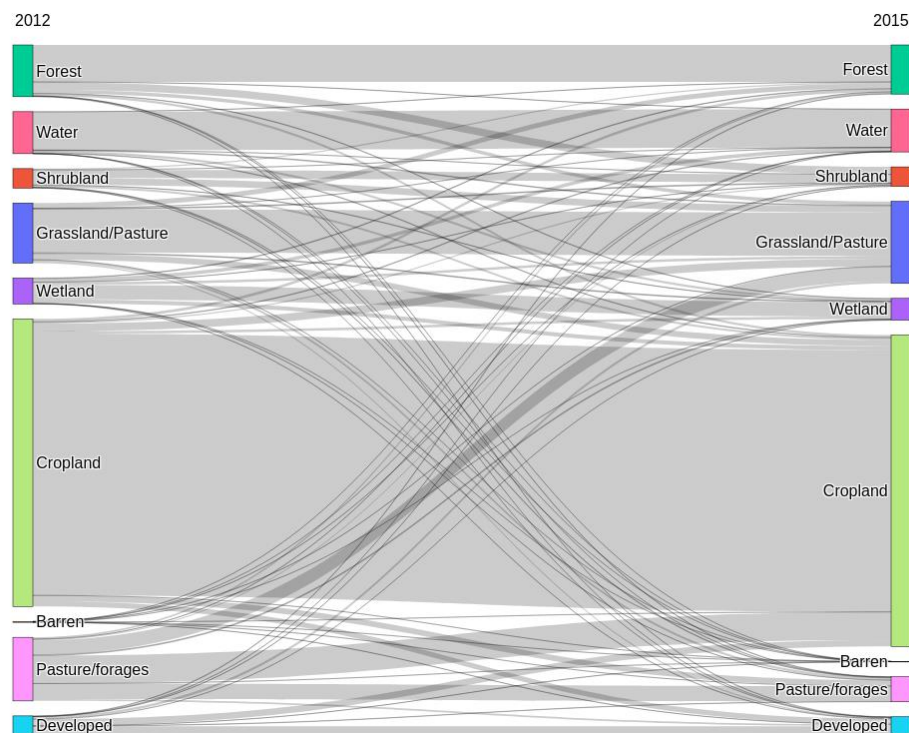


Figure 2.1. Sankey diagram of land use and land cover (LULC) changes between 2012 and 2015 using annual crop inventory (ACI) data.

Grassland monitoring using remote sensing technology has gained significant attention in recent years due to the spatial representation of landscape heterogeneity and the temporal frequency of monitoring growth dynamics (Zhou et al., 2024). Continuous assessments provide a unique opportunity to quantify grassland changes, distinguish between grassland types (Barrett et al., 2014), evaluate grazing impacts (Huo et al., 2024; Stumpf et al., 2020),

promote conservation and pasture stewardship (Lyu et al., 2020), and measure disturbances caused by drought and wildfire (Liu et al., 2019; Yan et al., 2020; Zhang et al., 2018).

Remote sensing methods are beneficial because they offer a non-destructive method of collecting data from extensive geographic areas repetitively and frequently over time (Inoue, 2003). In-situ observations are limited in monitoring large areas due to time and financial resources, and remote sensing has been increasingly used to address these problems (Muro et al., 2022; Zhang et al., 2022). Lyu et al. (2024) reviewed the use of remote sensing technology for grassland biodiversity monitoring in relevant studies from 2000 to 2022. They found out that current monitoring of grassland biodiversity has faced some problems, including the lack of fusing multisource remote sensing and a suitable analytical tool. They recommended combining the air and ground sensor networks to regularly observe grassland biodiversity in different ways, over a range of areas and throughout the food chain (Lyu et al., 2024). Remote sensing data can be used to measure various ecophysiological variables of a plant, such as biomass, leaf area index (LAI), phenology, yield, transpiration, photosynthesis, physiological stress response, water, chlorophyll and nitrogen contents, at a large scale (Inoue, 2003).

Multispectral spaceborne sensors (MSS) have been the primary data source for mapping grasslands (X. Wang et al., 2022). Landsat satellite series images with 30 m spatial resolution provide a great opportunity for long time-series analysis with datasets available since the 70's (Reinermann et al., 2020; Schwieder et al., 2022). The Moderate Resolution Imaging Spectroradiometer (MODIS) has a continuous (i.e., more frequent revisits) regional coverage with various spectral and spatial resolutions since 2002 (Hubert-Moy et al., 2019; Paltsyn et al., 2017). Since 2017, imagery collected by Sentinel-2 constellations at 10 m spatial resolution has also been used for grassland classification and mapping its spatial distribution (Abdollahi et al., 2022; Badreldin et al., 2021). Among multispectral remote sensors, MODIS, Landsat and Sentinel are the three most commonly used in grassland applications (Z. Wang et al., 2022). Recently, there has been a tendency to use sensors mounted on drones in monitoring grasslands. These unmanned aerial vehicle (UAV) sensors, particularly useful for monitoring grassland at local and farm scales, are beneficial in agriculture, mapping, and other fields (Zhang et al., 2022). If they combine with hyperspectral sensors, their high-spatial

resolution and high-spectral resolution make them suitable for monitoring grassland degradation (Zhang et al., 2022).

Spectral vegetation indices (VIs) derived from MSS products have been frequently used to discriminate between different terrestrial ecosystem species (Ali et al., 2016). Normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) are two common VIs used to monitor grasslands (Paltsyn et al., 2017). McInnes et al. (2015) have used seasonal NDVI time series of MODIS to classify grasslands in the grasslands Prairie region of Alberta with 73% overall accuracy (OA) in the dry mixed normalized difference moisture index (NDMI). Frank et al. (2022) demonstrated the capability to use three vegetation indices (VIs), namely NDVI, NDMI, and plant senescence reflectance index (PSRI), to differentiate among six grassland/forage types in three Canadian ecoregions located in Saskatchewan. Their classification results indicated that VIs can separate various classes in the Moist and Mixed Ecoregions, with some challenges in the Fescue Ecoregion (Frank et al., 2022). Badreldin et al. (2021) mapped the distribution of native, tame and mixed grasslands using massive remote sensing data and various vegetation features across Saskatchewan's mixed grassland ecoregion with an OA of 90.2%, which indicated the importance of developing a classification strategy that considers landscape heterogeneity (e.g., soil type and climate regime).

Besides optical remote sensing, radar remote sensing has also been utilized to map grasslands, either individually or in combination with optical sensors. Unlike optical remote sensing, synthetic aperture radar (SAR) remote sensing is not restricted by atmospheric interventions, such as clouds and aerosols that limit the short wavelength transmission (Molema et al., 2025). Although radar imagery is less visually interpretable than optical imagery, it can penetrate objects and provide valuable information about the vertical structure of vegetation (Canada Centre for Remote Sensing, 2010). SAR data can enhance land cover classification, especially when combined with optical data (Prasad et al., 2022). Prasad et al. (2022) reported that integration of Sentinel-1 and Sentinel-2 made a high OA ranging from 87.0 % to 89.5 % over different years. Fatholouloumi et al. (2022) evaluated the effect of surface biophysical characteristics, such as moisture, greenness and surface backscattering, derived from optical and radar sensors on land crop/cover identification and classification in

Ontario, Canada. Their results also showed that the combination of features achieved better classification performance in land cover classification.

With the increased availability of remote sensing products, the importance of assessing and using supervised machine learning algorithms has increased dramatically to assess their use in a wide range of applications, such as monitoring of land cover conditions and dynamics, as well as detailed landscape classification (Abdi, 2020; L. N. Smith et al., 2019; Willcock et al., 2018). Large, complex datasets have been effectively processed by various supervised and unsupervised machine learning algorithms for large-scale land cover classification (Yu et al., 2024). For example, Zhao et al. (Zhao et al., 2022) classified grasslands in Zambia using the random forest (RF) type of machine learning. Similarly, B. Sun et al. (2024) also used RF for grassland classification, indicating that RF is a reliable grassland classification method. Other types of popular machine learning methods, like support vector machine (SVM) and artificial neural networks (ANN), were also used for grassland classification (N. Wu et al., 2023). Recently, researchers have been leveraging deep learning technology like convolutional neural networks to extract high-level image features, improving learning efficiency (Y. Wang et al., 2023).

This study aimed to classify native, tame, and mixed grassland classes across approximately 69,647 km² of Manitoba's PE using the Google Earth Engine (GEE), a cloud-based platform that facilitates the analysis and visualization of large-scale geospatial datasets. The methodology consisted of a novel classification strategy designed to obtain the highest grassland classification accuracy in highly diverse ecoregions and soil types.

A three-step classification scheme was designed to create a reliable grassland map in the final step. The first two steps primarily focused on distinguishing grassland from non-grassland areas. This classification process utilized satellite data (optical and radar), topographic data generated by the Shuttle Radar Topography Mission (SRTM), and vegetation indices (VIs) extracted from optical data as input for three machine learning algorithms, namely, RF, SVM and gradient tree boosting (GTB). The research goals were as follows: (1) to determine the optimal satellite data combination among Sentinel-2, Landsat 8 and Landsat 9 and combined with SAR data such as Sentinel-1; (2) to evaluate the performance of three widely used machine learning algorithms, such as RF, SVM, and GTB, in mapping grassland through three steps. A feature selection technique was applied to improve classification

accuracy and reduce dimensionality; (3) to investigate the effect of image fusion techniques on grassland classification by using combined data from multiple sensors to enhance data quality beyond what a single data source can provide. Two main fusion techniques using hue, saturation, and value (HSV) and the Brovey transform (BT) were included in this study. For objectives 1 and 2, multi-temporal data were used to increase the number of variables and account for phenological growth stages in the models. Lastly, a web map was developed to show the results of all tests conducted in the context of the Soil Landscapes of Canada (SLC) database.

2.3. Materials and Methods

2.3.1. Study Area Description

The study zone spanned across Manitoba's PE with a total area of 69,647 km² (Figure 2.2). The area includes different ecoregions, such as the Aspen Parkland, the Lake Manitoba Plain and the Southwest Manitoba Uplands (Encabo et al., 2023). A large proportion of the region is used for agriculture, with smaller areas of natural land cover. Woodlands comprise a small percentage of the land cover in PE, mainly in the Aspen Parkland ecoregion. The grassland species in Manitoba's PE can be found in all ecoregions, consisting of native species like needle grasses, June-grass, wheat-grasses, blue grama grass, pasture sage, and non-vascular species. Additionally, there are tame species such as crested wheat-grass, alfalfa, Russian wild rye, and sweet clover. Precipitation varies from less than 280 mm/yr in the dry core area to 540 mm/yr in the Lake Manitoba Plain Ecoregion. The climate is variable, with cool to very cold winters (average of -6 to -17 °C), and warm, moist summers (average of 15 to 19 °C) (R. E. Smith et al., 1998). In Lake Manitoba Plain and Aspen Parkland ecoregions, Black Chernozemic soils are the dominant soil type, while in Southwest Manitoba Upland Dark Gray and Black Chernozemic soils are the primary types (R. E. Smith et al., 1998). The climate moisture index (CMI), determined by subtracting potential evapotranspiration from the total annual precipitation, showed a water deficit ranging between 200 and 250 mm for this area (Anderson & Cerkowniak, 2010).

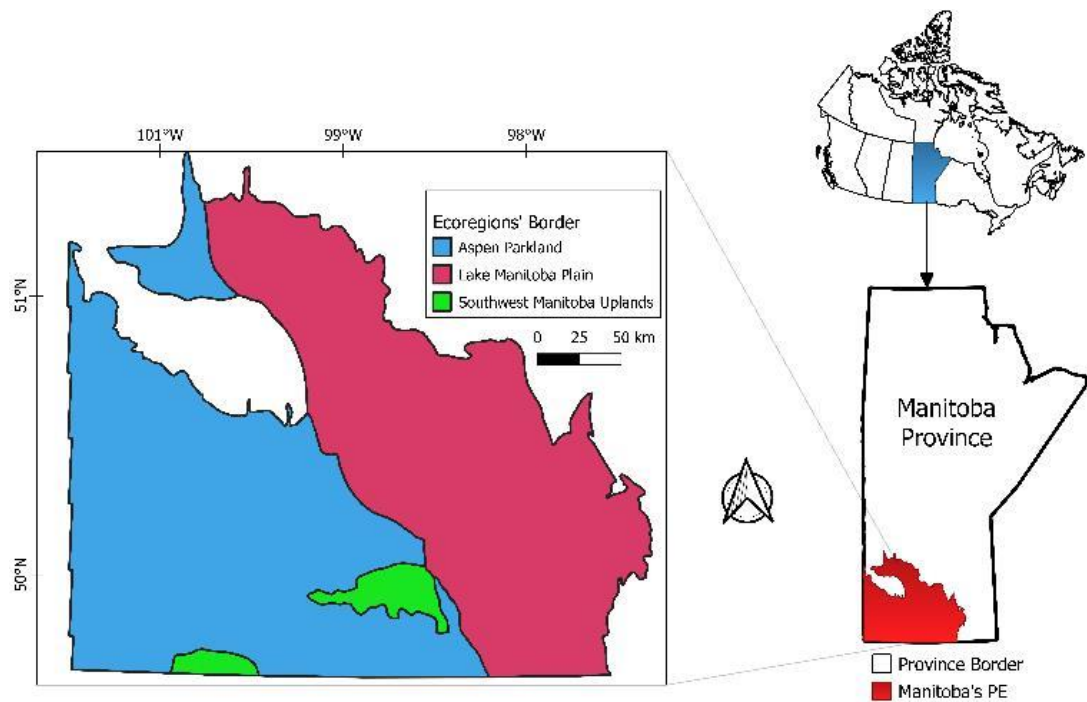


Figure 2.2. Geographical location of the study area. Manitoba's PE with different ecoregions is located in Manitoba Province, Canada.

2.3.2. Field Survey Sampling

This research used the comprehensive description of grassland mapping classes developed by Badreldin et al. (2021), which considers the MSS and SAR data limits, such as temporal and spatial resolutions. Precise definitions for tame, native, and mixed grasslands (Table 2.1) were established to guide the field survey team in collecting accurate and detailed field data. To identify the three main grassland types, 3,004 field points were obtained from the extensive ground-truth efforts carried out by the Canadian Forage and Grassland Association (CFGa) and the Ministry of Agriculture in Manitoba between 2021 and 2023 (Figure 2.3). The large number of ground samples strengthened the robustness and accuracy of the supervised machine learning classification models.

Table 2.1. The definition for grassland and other land use and land cover classes, adapted from Badreldin et al. (2021)

Class Name	Class Definition	Landscape Condition
Tame	Predominantly native grasses such as: • crested wheat-grass • smooth brome • Russian wild rye • alfalfa • sweet clover	If $\geq 75\%$ of the sample area is made up of planted or introduced grasses and forbs.
Native	Periodically cultivated, such as: • Needle grasses (needle and thread, porcupine grass, and green needle grass) • Wheat-grasses (slender wheat-grass, western wheat-grass, and awned wheat-grass) • June-grass, blue grama grass, sedge species, pasture sage, and non-vascular species (selaginella or lichens)	If $\geq 75\%$ of the sample area is made up of native grass species
Mixed	A heterogeneous area in which neither native nor tame grasses	If 50% have the majority, and the sample area was combined with other classes, such as bare soil or unidentified invasive species
Cropland	Oilseeds, Pulses, Grain, and Potatoes. (e.g., barley, oats, rye, spelt, wheat, corn, canola, soybean, and peas)	
Woodland	Predominantly forested or treed areas such as coniferous, broadleaf and mixedwood; shrub vegetation	
Water bodies	Lakes, Reservoirs, Rivers, Streams, and Wetlands (e.g., Bog, Fen, Marsh, and Swamp)	
Built-up	Roads and other man-made structures such as railroads, Buildings include both residential and industrial buildings, industrial sites, mine structures, lands with exposed soil, sand, or rocks	

Additional land cover and land use classes were introduced to improve the accuracy of the supervised machine learning classification approaches used. These supplementary classes were extracted from the AAFC ACI and the European Space Agency (ESA) WorldCover products, with spatial resolutions of 30 m and 10 m, respectively (AAFC Canada, 2013; Zanaga et al., 2022). The detailed land cover classes were aggregated into broader categories (Table 2.1); for example, the coniferous, broadleaf, mixedwood, and shrub classes were merged into the woodland class. This aggregation strategy was crucial for improving the accuracy of grassland mapping and for enabling the machine learning algorithm to distinguish between grassland and non-grassland classes effectively. A total of 7,570 samples were used for training and testing across all land cover classes (Figure 2.3), including 3,004 grassland

samples (with tame accounting for 37.73%, native for 30.73%, and mixed for 31.79% of the samples) and 4,566 non-grassland samples (with cropland comprising 62.59%, water bodies 3.99%, woodland 21.70%, and built-up 11.72% of the samples). Overall, 90.53% of the ground samples represented vegetative classes such as cropland, native, tame, mixed, and woodland, while 9.47% represented non-vegetative classes like water bodies and built-up areas.

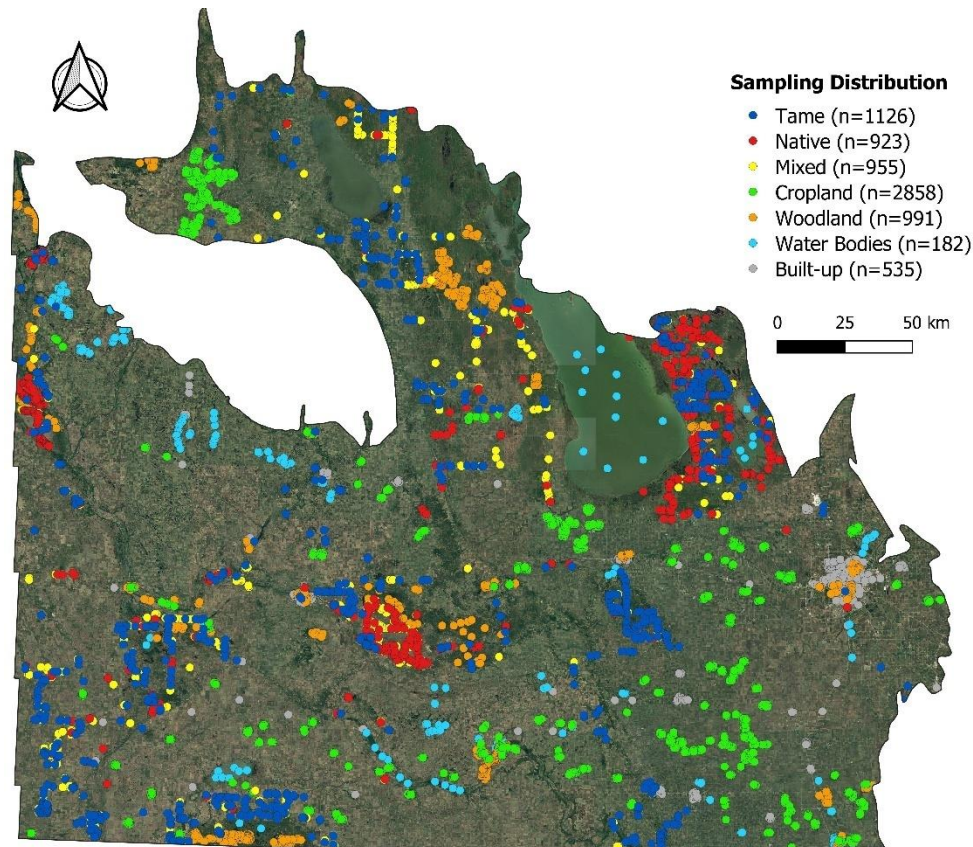


Figure 2.3. The spatial distribution of the ground-truthing sampling for all land cover and land use classes included in the classification of Manitoba's PE grasslands.

2.3.3. Grassland Classification Strategy

In this study, a novel strategy for grassland classification has been developed to enhance the accuracy and reliability of grassland mapping in Manitoba's diverse ecosystems, which was based on previous recommendations (De Fioravante et al., 2021; Yu et al., 2024). This strategy integrates multi-source remote sensing data (e.g., Sentinel-1, Sentinel-2, Landsat 8, and Landsat 9) with advanced machine learning techniques; this study used three supervised machine learning techniques (RF, SVM, and GTB). Unlike previous approaches, this strategy uses a three-step hierarchical classification process that incrementally refines the classification from vegetation vs. non-vegetation (step 1), grassland vs. non-grassland (step

2), leading to the final stage of specific grassland classes (step 3) to identify three grassland types (native, mixed, and tame).

As shown in Figure 2.4, the workflow underpinning this strategy details the data preprocessing, feature selection, and classification steps. The workflow operationalizes the strategy by systematically incorporating ancillary data, such as topographic features and field survey results, to improve model performance. This alignment between the workflow and the classification strategy ensures that the methodology is robust to landscape heterogeneity and applicable across various grassland types.

Image fusion was performed by combining Sentinel-2 and Landsat datasets using HSV and BT techniques. This approach aimed to enhance spatial and spectral resolution to improve grassland classification accuracy. The rationale for exploring image fusion was to provide an alternative method for data enhancement that could achieve reliable classification results without relying on multi-temporal data composition, which demands greater computational resources. However, this complements rather than replaces the multi-temporal composition strategy within the overall workflow, which remains critical for capturing seasonal phenological growth stages and improving classification accuracy.

Eighty-nine input variables, including topographic features (e.g., elevation, slope, and aspect), spectral bands, SAR data, VIs including NDVI, PSRI, NDMI, ratio vegetation index (RVI), renormalized difference vegetation index (RDVI), EVI and Green Normalized Difference Vegetation Index (GNDVI), and AAFC/ESA land cover data, were used in the analyses. These datasets were used during the three steps of grassland classification (Figure 2.4), which consisted of distinguishing between vegetated and non-vegetated pixels (step 1), discriminating grassland and non-grassland classes (step 2), and classifying native, tame and mixed grassland classes (step 3). The SLC version 3.2 shapefile was used to summarize the area of each grassland class and landscape type within each ecoregion. Zonal statistics were used to calculate the area and percentage for a final web-map visualization.

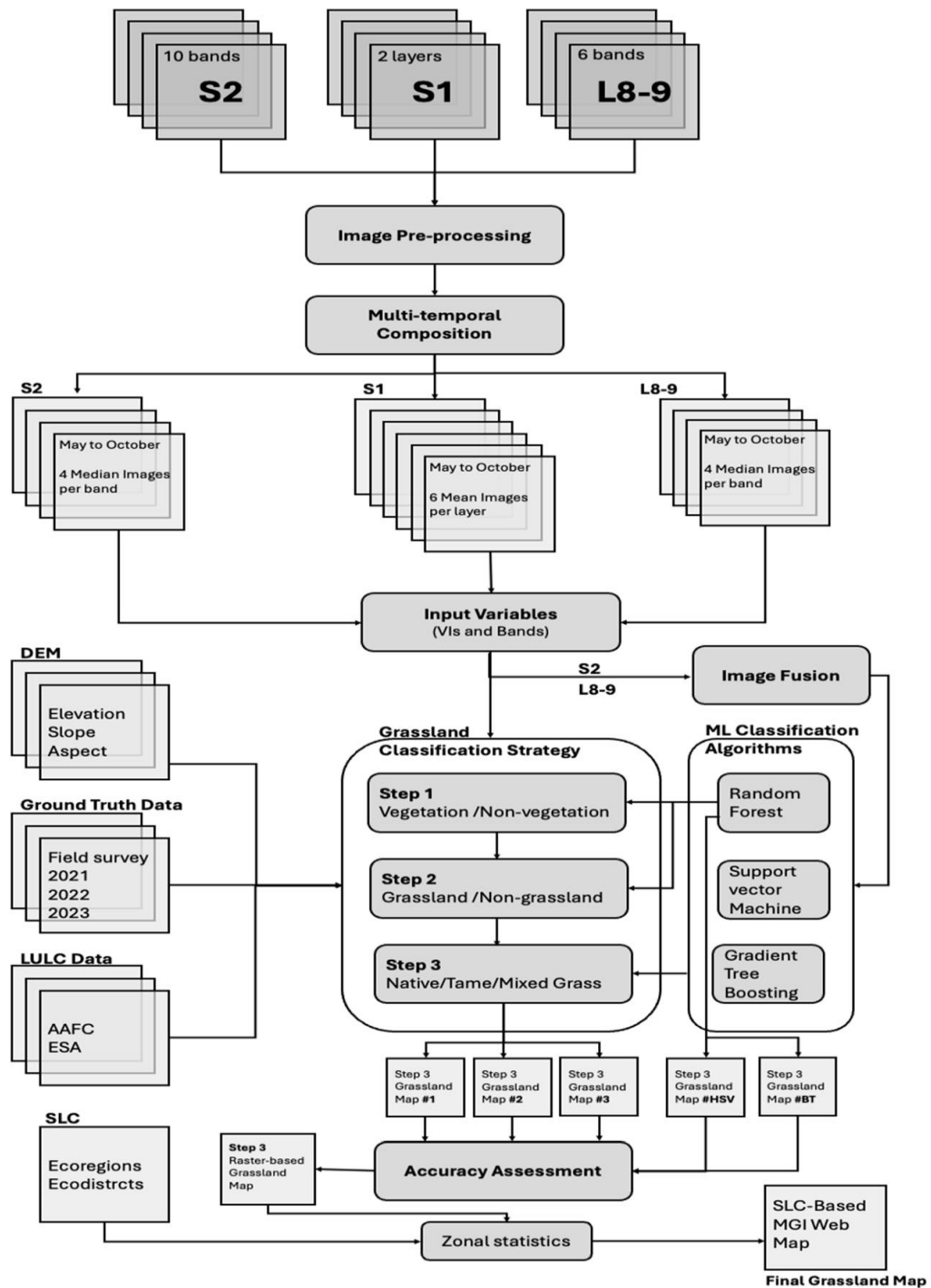


Figure 2.4. Workflow of the novel grassland classification strategy developed for enhanced ecological monitoring using multi-source remote sensing data and advanced machine learning techniques. The workflow integrates data from Sentinel-1, Sentinel-2, Landsat 8, and Landsat 9 satellites. It involves stages of image pre-processing, multi-temporal composition, image fusion using HSV and BT, and advanced machine learning classifiers, including RF, SVM, and GTB. The classification is performed in three steps to achieve fine-scale identification of native, tame, and mixed grasses, starting from basic vegetation classification (Level 1) to detailed grassland class differentiation (Level 3). Ancillary field data, topographic features, and LULC information were incorporated as inputs to generate the final grassland maps for web-based visualization.

2.3.4. Satellite Data Management

The ESA operates Sentinel satellites that collect SAR (Sentinel-1 satellites) and multispectral (Sentinel-2 satellites) data for Earth monitoring (Drusch et al., 2012a; Torres et al., 2012). Sentinel-1 comprises a constellation of two polar-orbiting satellites collecting C-band SAR imagery, which is not affected by weather conditions (Potin et al., 2016). Sentinel-1 imagery on GEE consists of level-1 ground range detected (GRD) scenes processed to backscatter coefficient (σ^0) in decibels (dB). Sentinel-2 comprises a constellation of two polar-orbiting satellites, covering 13 spectral bands from visible and near-infrared to short-wave infrared (Main-Knorn et al., 2017).

Landsat satellites are operated by NASA (National Aeronautics and Space Administration) and have been continuously observing Earth since 1972 (Williams et al., 2006; Wulder et al., 2022); Landsat 8 and Landsat 9 are two operational satellites from this series. Landsat 9 was launched in 2021 and replaced Landsat 7. The combined operation of Landsat 8 and 9 (Landsat 8-9) decreased revisit time to 8 days (Markham et al., 2020). Landsat data are moderated by the USGS (United States Geological Survey). In this study, Landsat images were used to compare the classification accuracy for different sensor combinations in step 1 and step 2 classification, as well as in fusion. Some studies, such as Yu et al. (2024), emphasized that topographic features increased the accuracy of grassland classification. Thus, the digital elevation model (DEM), slope and aspect, extracted from SRTM data, were used as auxiliary data to improve the classification accuracy. The major characteristics of all the remote sensing datasets used in the analyses are provided in Table 2.2.

Table 2.2. An overview of satellite datasets used in this study summarizes the satellite sensors employed in the study and highlights the acquisition dates, spatial resolution, and specific rationale for each sensor's inclusion in this study.

Sensors	Date of Data Acquisition	Spatial Resolution	Rationale of usage	References
Sentinel-1	01/05/2023 -31/10/2023	10 m	• Cloud-penetrating radar capability ensures consistent data collection even in cloud-covered regions. • Provides valuable microwave information for assessing vegetation structure and moisture content, aiding in the differentiation of grassland types.	(Torres et al., 2012)
Sentinel-2	01/05/2023 -31/10/2023	10 m	• Offers the highest spatial resolution available for free, enabling detailed grassland mapping.	(Drusch et al., 2012b)

			• Wide spectral range, including vegetation-specific bands, allows for precise discrimination of grassland types and health assessments.	
Landsat 8 Collection 2 Tier 1 Surface Reflectance	01/05/2023 -31/10/2023	30 m	• Long-term data record supports temporal comparisons and change detection. • Utilized for spatial and spectral consistency evaluation when compared to Sentinel-2, and aids in data fusion for enhanced mapping accuracy.	(Roy et al., 2014)
Landsat 9 Collection 2 Tier 1 Surface Reflectance	01/05/2023 -31/10/2023	30 m	• Continuation of the Landsat program with improved radiometric resolution. • Used for cross-sensor comparison and fusion with Sentinel-2 data, ensuring a robust dataset for grassland classification.	(Masek et al., 2020)
SRTM	11/02/2000 -22/02/2000	1 Arc (~30 m)	• Provides crucial topographic information that enhances land cover classification accuracy. • Incorporates elevation data, important for understanding how topography influences vegetation patterns in grasslands.	(Farr et al., 2007)

Both raw and preprocessed data are available in GEE. In this study, the level-2 data of Sentinel-2 and Landsat as preprocessed data were used. Image preprocessing increases spectral quality because MSS data are subject to distortion due to sensor characteristics, solar radiation, atmospheric conditions, and topography (Young et al., 2017). Atmospheric correction is required for several landscape modelling applications, such as land cover classification and change detection (Song et al., 2001). The study selected atmospherically corrected datasets (Surface Reflectance level data) to minimize the influence of atmospheric conditions, which is especially important given the large study area and varying data acquisition times. Clouds and their shadows in MSS data can also hinder their utilization for analysis (Zhu & Woodcock, 2012), as shown in Figure 2.5. For Sentinel-2 images, the Cloud Score+ dataset was utilized to identify relatively clear pixels and effectively remove clouds and cloud shadows from L2A (Surface Reflectance) imagery (Earth, 2024). The Cloud Score+ dataset includes two quality assessment (QA) bands. For Sentinel-1 imagery, preprocessing steps, including thermal noise removal, terrain correction, and GRD border noise removal, (Filipponi, 2019), were applied to Sentinel-1 imagery in GEE to derive the backscatter coefficient in each pixel. For Landsat 8 and Landsat 9 images, the QA ("QA_PIXEL") band was used with a bit mask technique to remove clouds and shadow pixels. This preprocessing step

was performed on the level-2 dataset of the validated Tier 1, known for its high geometric and radiometric quality (Landsat Algorithms | Google Earth Engine, n.d.).

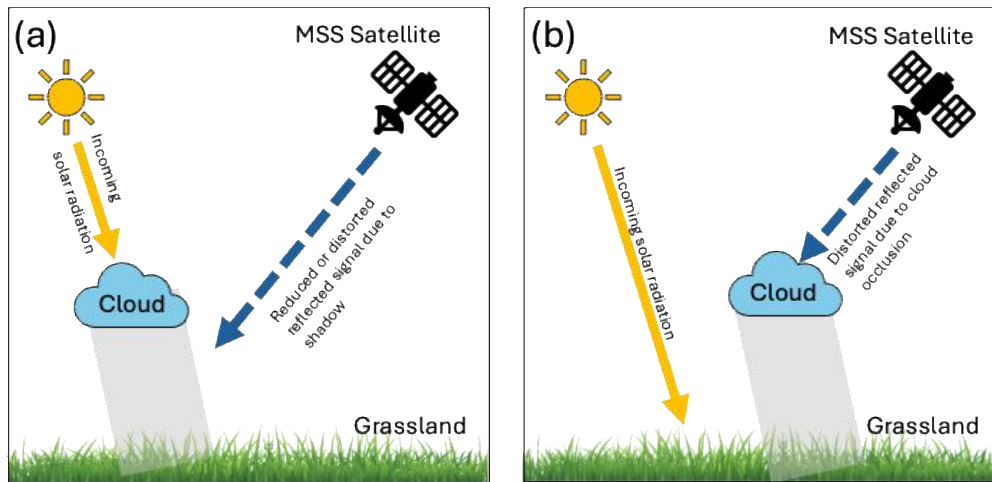


Figure 2.5. Atmospheric effects on remote sensing imagery demonstrate the influence of atmospheric conditions on the quality of remote sensing data, specifically how clouds and shadows can occlude pixels and impact the accuracy of the reflected signal received by MSS sensors; (a) Pixel occluded by shadow shows a scenario where a shadow, cast by an obstacle like a cloud, causes the pixel to be occluded, leading to distorted signals received by the sensor; and (b) Pixel occluded by cloud shows a situation where a cloud directly occludes the pixel, resulting in inaccurate data due to the cloud's interference with the reflected sunlight that reaches the sensor.

During the active growing season of grasslands in PE, which spans from May to October, multi-temporal compositions were used to distinguish between different types of grassland effectively. Due to the extensive study area and cloud masking, creating a mosaic of the entire area using images taken within a month was challenging. Thus, a minimum period of four weeks was considered to generate a mosaic from images acquired throughout the period (Figure 2.6). For temporal aggregation, the Median function in GEE was used to reduce the amount of data and decrease the remaining cloudy pixels in the final mosaic image (Figure 2.7). This process was repeated during the selected growing season to create a stacked image for further image processing. For example, in Sentinel-2 data, stacked images were made for the four periods of the growing season, each without overlap. The same process was also applied to Landsat data used in step 1 and step 2 classification. In addition, the temporal resolution was reduced by merging Landsat 8 and Landsat 9.

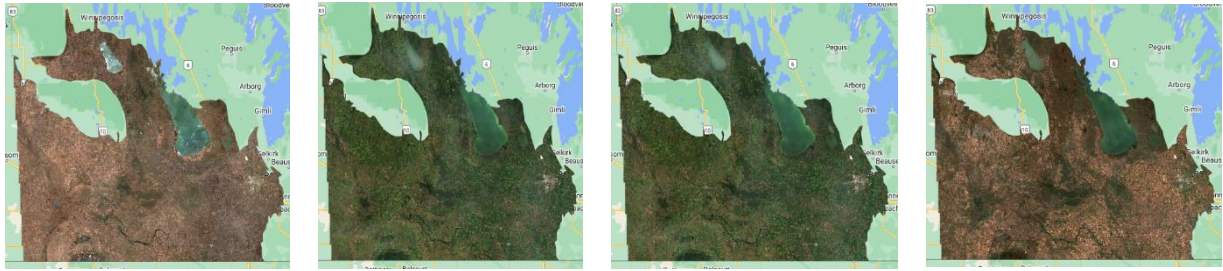


Figure 2.6. Multi-temporal mosaic of the study area during the growing season period using Sentinel-2 images.

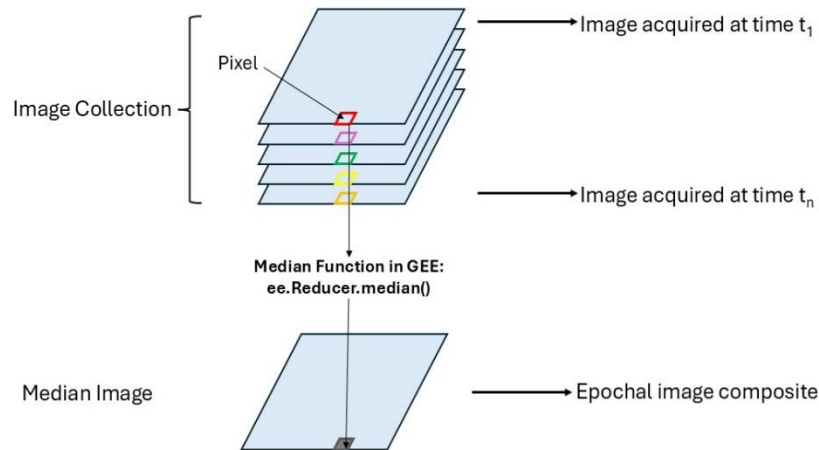


Figure 2.7. Median function used in Google Earth Engine (GEE) to reduce an image collection.

Given that Sentinel-1 data quality is robust to atmospheric disturbances, it was possible to generate a monthly mosaic image of the entire study area each month. The Mean function in GEE was utilized to minimize the effects of speckle noise when aggregating the multi-temporal SAR data (Argenti et al., 2013). The stacked image comprising six monthly mean images was created for Sentinel-1.

The preprocessed images resulted in four groups of raster data used as input variables for supervised machine learning classification algorithms (Table 2.3). These inputs included surface reflectance of spectral bands [i.e., Blue, Green, Red, NIR, SWIR 1, SWIR 2 and Red Edge 1-4 (Only in Sentinel-1)]; VIs [e.g., NDVI, PSRI, NDMI, RVI, RDVI, GNDVI and EVI]; topographic variables [i.e., elevation, slope, aspect]; and SAR data [i.e., VH, VV polarization and their ratio].

Table 2.3. Description of all variables obtained through several sensors and utilized in this study.

Dataset group type	Input Variable
Spectral Bands	Sentinel-2:
	Blue; Green; Red; NIR; SWIR 1& 2; Red Edge 1-4

	Landsat 8 and Landsat 9: Blue; Green; Red; NIR; SWIR 1 and 2
SAR Variables	Sentinel-1: VV; VH; VH/VV
Vis	NDVI (Rouse et al., 1974): $(\text{NIR} - \text{Red})/(\text{NIR} + \text{Red})$ PSRI* (Merzlyak et al., 1999): $(\text{Red} - \text{Blue})/\text{Red Edge}$ NDMI (Wilson & Sader, 2002): $(\text{NIR} - \text{SWIR1})/(\text{NIR} + \text{SWIR2})$ RVI (Birth & McVey, 1968): $(\text{NIR})/(\text{Red})$ RDVI (Roujean & Breon, 1995): $(\text{NIR} - \text{Red})/(\sqrt{\text{NIR} + \text{Red}})$ GNDVI (Gitelson et al., 1996): $(\text{NIR} - \text{Green})/(\text{NIR} + \text{Green})$ EVI (Huete et al., 1994): $2.5 \times (\text{NIR} - \text{Red})/(\text{NIR} + 6 \times \text{Red} - 7.5 \times \text{Blue} + 1)$
Topographic Variables	SRTM: Elevation; Slope; Aspect

*As Sentinel-2 has dedicated red-edge bands (Bands 5, 6, 7), this index was used exclusively with Sentinel-2. B6 was used as Red Edge (IDB)

2.3.5. The Supervised Machine Learning Classification Models

Grassland classification was performed using RF, SVM, and GTB algorithms on GEE in three steps. In steps 1 and 2, different combinations of Sentinel-1 SAR, Sentinel-2 and Landsat were tested using the RF method to find the best sensor combination according to classification accuracy. Step 1 distinguished between vegetation and non-vegetation classes, while step 2 identified grassland and non-grassland classes. Subsequently, when grassland class was identified, three widely used classifiers [i.e., RF, SVM and GTB] were used to classify the three distinct grassland classes [i.e., native, tame, and mixed].

RF method belongs to the ensemble learning methods (Breiman, 2001) and combines multiple decision trees (DTs) to produce a more accurate and robust result than any of the individual DTs alone. RFs build multiple DTs during training. Each tree is trained on a random subset of the training data and a random subset of the features (Shaik & Srinivasan, 2019). This randomness helps to ensure that the individual trees are diverse and not overly correlated, reducing the risk of overfitting. During classification, each tree in the RF independently predicts the class label, and the final prediction is determined by a majority vote (for classification tasks) or averaging (for regression tasks) among all the trees (M et al., 2023).

If there are T trees in the RF model, and each tree t predicts a class $P(x)_t$. The final prediction is the majority of those predictions with the highest votes [Equation (1)].

$$P(x) = \arg \max_{c \in C} (\sum_{t=1}^T \delta(P(x)_t, c)) \quad (1)$$

Where $P(x)$ is the predicted class for input x , c is the set of possible classes, T is the total number of DTs in the RF, $P(x)_t$ is the class predicted by the t^{th} DT for input x , $\delta(P(x)_t, c)$ is an indicator function that equals 1 if $P(x)_t = c$ and 0 otherwise.

The two most critical input factors for RF are the number of features (*mtry*) per split and the number of trees (*ntree*), both of which are specified by the user. Previous studies suggested that the optimal number of trees typically ranges from 100 to 500, while the optimal number of variables is often the square root of the total number of variables (Loukika et al., 2021).

Similar to RF, SVM is a supervised machine learning algorithm often used for image classification. SVM methods try to find an ideal hyperplane that effectively separates different sample sets. A hyperplane is a plane in a high-dimensional space that maximizes distance from all samples. In cases where the dataset cannot be linearly separated, two techniques, namely soft margins and the kernel trick, can be used. Soft margins allow the algorithm to misclassify a few elements while still preserving the algorithm's predictive capacity. The kernel trick uses a kernel function that maps data into higher-dimensional spaces, where class separation may be achieved. Several popular kernels commonly used in SVM are linear [Equation (2)], polynomial [Equation (3)] and radial basis function (RBF) [Equation (4)] (Abe, 2010). If the training dataset includes n points of the form $(x_1, y_1), \dots, (x_n, y_n)$, the different kernels can be defined as:

$$\text{Linear kernel: } k(x_i, x_j) = x_i \cdot x_j \quad (2)$$

$$\text{Polynomial kernel: } k(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d, \gamma > 0 \quad (3)$$

$$\text{RBF kernel: } k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (4)$$

The mathematics behind SVM algorithms was described in detail by Burges (Burges, 1998).

The GTB model, introduced by Friedman (2001), integrates DT and boosting algorithms. Gradient boosting is iterative and involves sequentially combining weak learners, typically DTs, such that each new tree rectifies the errors of the previous one. Each iteration invokes the following formula [Equation (5)]:

$$F_{i+1} = F_i - f_i \quad (5)$$

Where F_i and f_i are the strong model and the weak model at step i , respectively.

The approach starts with a single DT and the performance of the tree is evaluated using a loss function (e.g. cross-entropy), as shown in Equation (6):

$$\text{Loss}(l, p) = -\sum l(x) \log p(x) \quad (6)$$

Where l is the labelled sample and p is the prediction. A loss value of 0 indicates perfect agreement between the prediction and labelled sample, whereas a loss of 1 signifies no agreement between them. Then, the second DT is added to the first one to lower its loss value. This process iterates until a stopping criterion is satisfied, such as reaching a maximum number of iterations or detecting overfitting of the model, as assessed on a separate validation dataset.

Tuning machine learning parameters is a complementary task that can improve classification accuracy and reduce computational cost by finding optimal parameters for each method (Z.-H. Zhou, 2021). Each model has several parameters that can be tuned. For instance, GTB and RF each have six parameters that can be optimized, while SVM is more complex, with multiple kernels, each having several parameters. Our objective was not to find the optimal parameters by testing a wide range of values for each parameter. Instead, the focus was on adjusting a few key parameters for the classifier models. In the RF model, *ntree* and the total *mtry* per split were selected using Out-of-Bag (OOB) error. For the SVM, the classification accuracy of three popular kernels—Linear, Polynomial, and RBF—was tested. In the GTB model, the effect of *ntree* on OA was investigated.

2.3.6. Dimensionality Reduction Technique for Big Data

Reducing the number of input variables is beneficial for cutting computational costs and sometimes boosting model performance (Kuhn & Johnson, 2013; Lv et al., 2022). Feature selection methods are divided into supervised and unsupervised methods. If a method

considers the outcomes when eliminating features, it is called supervised. Otherwise, the technique is considered unsupervised. Unsupervised methods try to remove the features that have high correlation with other features. But supervised methods try to select features that make models simple and increase the accuracy of the model (Kuhn & Johnson, 2013; Lv et al., 2022). Supervised methods fall into three categories: 1) filters, 2) wrappers, and 3) embedded (Pudjihartono et al., 2022). In this study, a supervised filter method was used to select features for analysis. Filter methods rank features based on the score they yield in different statistical tests. The analysis of feature selection was carried out using analysis of variance (ANOVA), which identified significant features by comparing the means of various groups. The F-statistic, as defined in Equation (7), was calculated in the ANOVA procedure to determine whether the differences in means between groups were statistically significant.

$$F = \frac{\text{Variability between groups}}{\text{Variability within groups}} \quad (7)$$

The features were ranked by F value and selected for further analysis (Pudjihartono et al., 2022).

Feature reduction using ANOVA involved selecting features that have a significant relationship with the target variable. Features were ranked based on their F-statistic values, where higher scores indicated greater discriminative power. This enabled us to decide which features are worth retaining for the classification task by RF, SVM and GT. In Google Colab, a feature selection process was implemented using the ANOVA F-value as the scoring function. Specifically, the `f_classif` function from `sklearn` was applied within a feature selector (`SelectKBest`) to evaluate features. The input for this process consists of `X`, which is the independent variables or features of the dataset, and `y`, which is the dependent variable or class labels for classification. For this study, `X` and `y` were extracted from the training dataset prepared for each classifier. The output was a list of features ranked in descending order based on their F-statistic scores.

To determine the optimal feature set, features were progressively added to the model, starting from those with the highest scores and moving toward those with lower scores. The overall accuracy (OA) of each classifier was evaluated at each step to identify the feature subset that yielded the best performance for grassland classification.

2.3.7. Classification Accuracy Assessment

Accuracy assessment was the final step of the process to find out how effectively the model classified the pixels. Validation (or testing) in this study was conducted using 30% of the total number of labelled samples. Actually, the validating part included those datasets that were not used to train the classifier; they represented the nominal or “real-life” accuracy of a map. Quantitative statements about the performance of classification models can be made using test accuracy. Over the past years, accuracy assessment has become more detailed and precise, and evolved from visual evaluation of the derived map to definition of sampling and response, and analysis of accuracy (Nicolau et al., 2024). The confusion matrix (Stehman & Foody, 2019) is one of the most used methods to characterize image classification accuracy (Foody, 2002). Several indices can be extracted from the confusion matrix, including OA, producer’s accuracy (PA), user’s accuracy (UA) and the Kappa coefficient, as defined in Equations (8-10). OA indicates what proportion of the test data was classified correctly. PA refers to the percentage of correctly classified pixels relative to the total number of classified pixels. UA represents the percentage of correctly classified pixels relative to the total ground truth data. The Kappa coefficient, ranging from -1 to 1, quantifies the agreement between observed and expected classifications. Higher values denote improved agreement and effectiveness of the classification model.

$$\text{Overall Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (8)$$

$$\text{Producer's accuracy (positive)} = \frac{TP}{(TP + FN)} \quad (9)$$

$$\text{User's accuracy (positive)} = \frac{TP}{(TP + FP)} \quad (10)$$

Where TP is defined as positive, and the actual class is positive, FP is classified as positive, and the actual class is negative, FN is classified as negative, and the actual class is positive, and TN is classified as negative, and the actual class is negative.

2.3.8. Image Fusion

Image fusion involves merging different images to create a new one that incorporates the best features from each. In the fusion process, an image with a higher spatial resolution and a multispectral image (MSI) with a lower spatial resolution are combined to build an MSI with a higher spatial resolution, which addresses the limitations of image sensors that cannot

simultaneously capture high-resolution and high spectral information (Kaplan, 2018). This process can be divided into three levels, depending on the stage at which fusion takes place: pixel level, feature level and decision level (Ghassemian, 2016). Pixel-level fusion methods (Figure 2.8) were utilized in this study to investigate the impact of image fusion on the classification algorithm at the step 3 classification stage across a subset of the study area. The BT and HSV approaches were applied to Landsat 8-9 and Sentinel-2 data. In the HSV method (H, S, and V stand for hue, saturation and value or intensity, respectively), the RGB bands of the image are transformed into the HSV colour space (Singh & Tiwari, 2021). Then, the V value of the image are transformed into the HSV colour space (Singh & Tiwari, 2021). Then, the V value of the converted HSV image was replaced with the DN value of a high-resolution (HR) image. This modified HSV image was converted back into an RGB colour image (Al-Wassai et al., 2011).

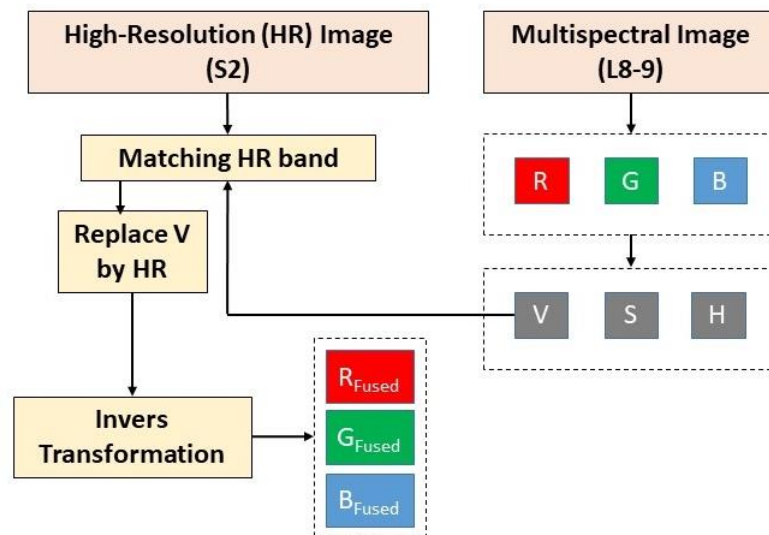


Figure 2.8. The steps of the HSV method, adapted from Al-Wassai et al. (Al-Wassai et al., 2011). After transforming the RGB image to HSV format, its V channel was replaced with the HR channel, which was then converted back to RGB mode.

In the BT method, the RGB bands of the MSI were proportionally combined with the HR image, and each of the multispectral bands was multiplied by the ratio of the HR image to the sum of the multispectral bands [Equation (11)]. This enhanced the spatial resolution while preserving the spectral characteristics of the original MSI, which resulted in a higher spatial resolution MSI that combines the details from both the multispectral and HR images (Mandhare et al., 2013).

$$FB = \frac{SP_{MSI}}{R_{MSI} + G_{MSI} + B_{MSI}} HR \quad (11)$$

Where SP_{MSI} refers to spectral bands of MSI (e.g., R_{MSI} , G_{MSI} , or B_{MSI}), FB represents fused bands corresponding to SP_{MSI} , and HR refers to the high-resolution band.

In both BT and HSV fusion techniques, the HR image used was the green band of Sentinel-2, while the Red, Green and Blue bands of Landsat, being low-resolution bands, were aimed at enhancing their resolution. Before fusing, Landsat 8 and Landsat 9 were initially merged, and their cloudy pixels were then masked. Similarly, cloudy pixels from Sentinel-2 were also masked. Finally, fusion results were evaluated both visually and quantitatively, using the coefficient of determination (R^2) as a quantitative metric.

2.4. Results

2.4.1. Accuracy Assessment for Steps 1 and 2

Achieving high classification accuracy at both step 1 (vegetation vs. non-vegetation) and step 2 (grassland vs. non-grassland) was essential for this novel grassland classification workflow to build a reliable, detailed grassland inventory. Accurate step 1 classification ensures that broad categories of vegetation cover, such as woodlands, croplands and grasslands, are correctly identified, which is crucial for subsequent, more detailed grassland analyses. Step 2 accuracy further refines this understanding by distinguishing between two different major vegetation types, such as grasslands, which are vital for improving grassland mapping in Canada; these accuracy assessments support the adopted approach in this research to test different data combinations between sensors and machine learning classification models (Figure 2.9). The highest classification accuracy was achieved by integrating data from both Sentinel-1 and Sentinel-2 (Sentinel-1+Sentinel-2), with PA and UA surpassing 97.5% for both vegetation and grassland classes (Table 2.4). The combination of multiple sensors enhanced the model's ability to discriminate between various land cover types accurately. The single-sensor approaches, although still effective, exhibited slightly lower accuracy, specifically for non-vegetation and non-grassland classes when Sentinel-1 is used alone. These findings imply that a more robust and accurate classification can be achieved by using multi-sensor satellite data, making it the preferred method for detailed grassland mapping in Canadian ecosystems.

The evaluation of classification accuracy for all steps in various machine learning models is considered crucial once the optimal sensor combination, such as Sentinel-1 + Sentinel-2, has been identified. The effectiveness of the classification process is dependent on the model's ability to accurately interpret and categorize sensor data, regardless of its quality. Discrepancies exist in the capabilities of different machine learning models, such as RF, SVM, and GTB, in terms of dealing with different data distributions and complexities. This research aims to determine the most precise and reliable model for classifying grasslands using the Sentinel-1 + Sentinel-2 sensor combination. Comparing the performance of the models is sought to minimize errors in subsequent analyses or decision-making processes. Understanding the disparities in model performance is intended to guide future studies and enhance the effectiveness of grassland mapping. Using Sentinel-1 + Sentinel-2 data and three distinct machine learning models [i.e., RF, SVM, and GTB], the study area was classified to distinguish grassland from non-grassland classes. As shown in Table 2.5, the RF model achieved the highest OA of 98.08% and a Kappa coefficient of 0.97, indicating the best classification agreement and precision. The GTB model, although highly accurate with an OA of 97.49% and a Kappa coefficient of 0.96, was found to perform slightly less effectively than the RF model. The SVM model attained an OA of 96.34% and a Kappa coefficient of 0.95, positioning its performance below that of RF and GTB. The performance of three classifiers was effective for the first two steps of grassland classification, with RF yielding the most accurate results.

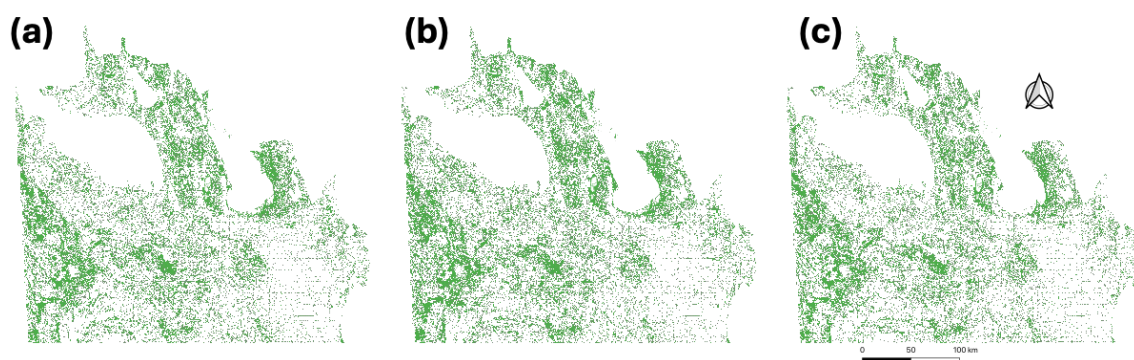


Figure 2.9. Comparison of step 2 grassland classification results using different machine learning models: (a) RF, (b) SVM, and (c) GTB.

Table 2.4. Classification Accuracy assessment of step 1 (Vegetation vs. Non-Vegetation) and step 2 (Grassland vs. Non-Grassland) classifications using the RF method for different sensor combinations (PA% - UA%).

Level \ Sensor type	Sentinel-1	Sentinel-2	Landsat 8-9	Sentinel1+ Landsat 8-9	Sentinel-1+ Sentinel-2
Step 1					
Vegetation class	94.86%-91.43%	96.79%-97.38%	97.03%-96.13%	96.73%-96.84%	98.59%-97.67%
Non-vegetation class	71.19%-94.73%	94.16%-94.99%	93.13%-96.62%	93.88%-95.18%	94.77%-99.14%
Step 2					
Grassland class	95.02%-90.01%	97.28%-96.55%	96.74%-97.53%	97.39%-97.50%	98.06%-98.18%
Non-grassland class	94.78%-92.14%	96.55%-97.79%	97.20%-95.43%	96.40%-96.51%	98.86%-97.41%

Table 2.5. The averaged OA and Kappa coefficient for different machine learning classification methods of the first two steps using Sentinel-1+Sentinel-2. The results show that DT-based classifiers have higher performance than SVM. However, RF achieved slightly higher accuracy than SVM in the first two steps.

Machine learning method	OA (%)	Kappa coefficient
RF	98.08	0.97
SVM	96.34	0.95
GTB	97.23	0.96

2.4.2. Accuracy Assessment for Step 3

Upon identification of the primary classes, a filtering process was initiated to exclude all classes other than grassland. This entailed the masking of all non-grassland pixels within the various data layers, encompassing Sentinel-2 spectral bands, VIs, Sentinel-1 data, and DEM. In step 3 classification, to prevent imbalanced classification problems (He & Garcia, 2009), the number of native grassland samples was oversampled using spatial techniques, such as buffers, to better balance with the number of tame samples. All variables that were masked were then used as input variables for the machine learning classification methods. As shown in Table 2.6, which compares the OA and Kappa coefficient for step 3 classification using RF, SVM, and GTB with and without feature selection techniques, several important insights emerge. The feature selection process has clearly contributed to improved classification performance across all methods. GTB, in particular, showed the greatest improvement in OA, increasing from 69.69% to 72.18%, and Kappa values from 0.55 to 0.58. RF demonstrated the largest feature reduction, decreasing from 89 to 61 features. SVM also benefited from feature selection, with SVM's OA improving from 56.58% to 60.54% and Kappa value from 0.36 to 0.40. These results suggest that feature selection is a valuable step in the detailed grassland classification process, particularly when dealing with complex datasets involving diverse landscapes with high biophysical similarity. The PA and UA of the tame,

native, and mixed classes for the RF, SVM, and GTB classifiers, with and without feature selection techniques, are shown in Figure 2.10. PA and UA of all grassland classes for RF and GTB increased after feature selection, but SVM showed some fluctuations. PA of the tame class and UA of native class in the SVM method experienced a slight decline, while other metrics improved after feature selection.

Table 2.6. Accuracy assessment of machine learning methods with/without feature selection technique for grassland classification using Sentinel-1+Sentinel-2 data.

Methods	Without a feature selection technique			With a feature selection technique		
	Number of features	OA (%)	Kappa	Number of features	OA (%)	Kappa
RF	89	69.96	0.55	61	72.62	0.58
SVM	89	56.58	0.36	67	60.54	0.40
GTB	89	67.69	0.50	76	72.18	0.58

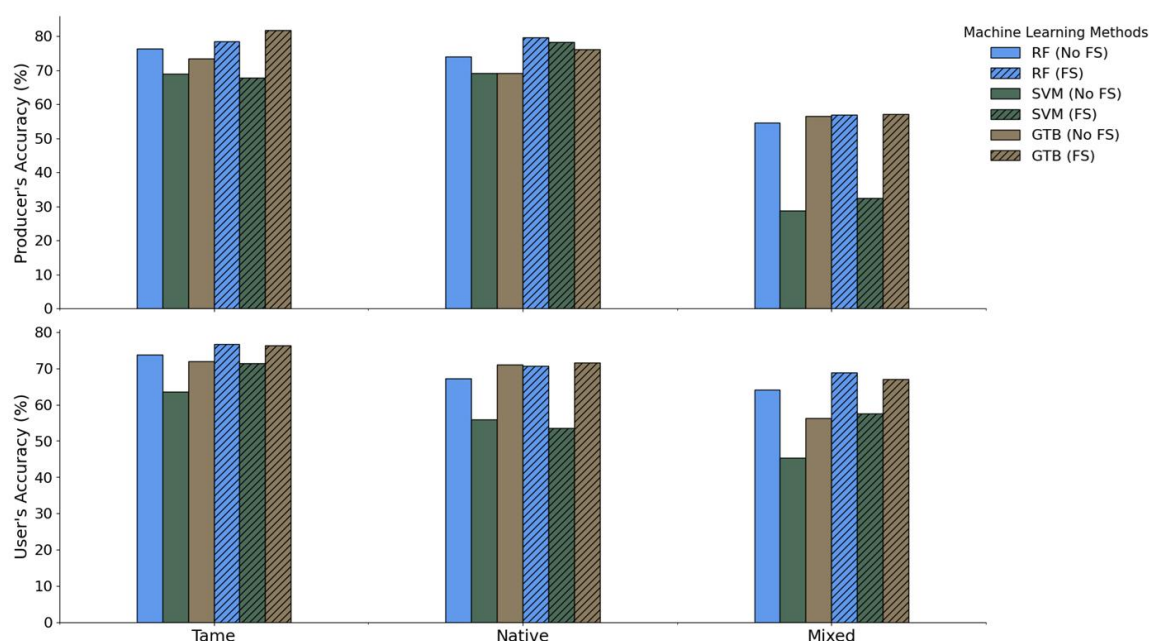


Figure 2.10. Producer's Accuracy (PA) and User's Accuracy (UA) of three machine learning models such as RF, GTB and SVM, with and without feature selection (FS).

As shown in Figure 2.11, the relationship between overall accuracy (OA) and the number of input features was depicted. The classification models were trained by sequentially incorporating features, beginning with the 10 most significant ones ranked based on ANOVA scores and progressively adding features in descending order of their statistical importance.

As the number of input features increased, the OA for all algorithms exhibited an initial rise before reaching a plateau. Figure 2.A1, in the Appendix of this chapter, demonstrates the score of each feature and indicates which features were excluded in each model when optimal OA was achieved.

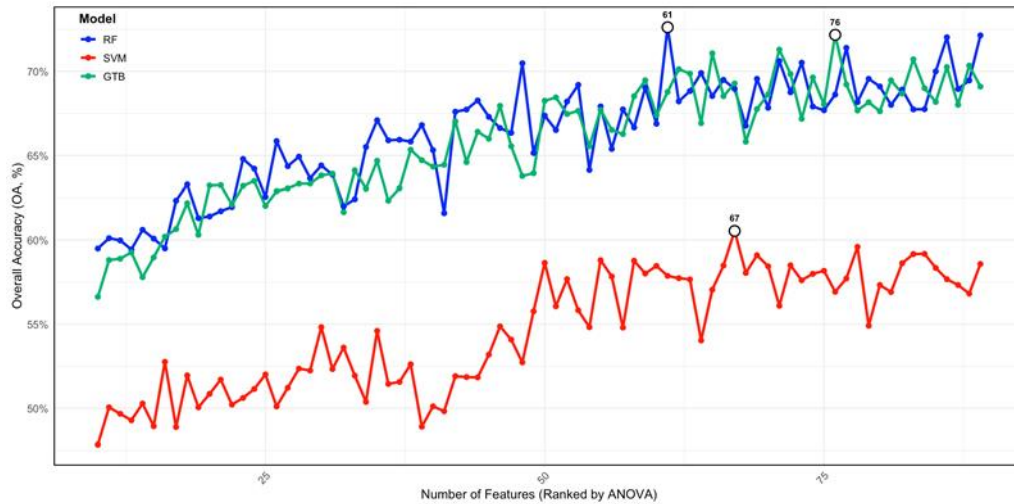


Figure 2.11. Classification accuracy varies with different input features ranked based on ANOVA for the RF, SVM, and GTB.

The established tuning parameters had significant impacts on the performances of RF, SVM and GTB classifiers. The results of hyperparameter tuning for different machine learning models are demonstrated in Figure 2.12-Figure 2.14. For the RF model, OOB error decreased more if the *mtry* per split was set to the square of *mtry* per split. The relationship between the *ntree* and OOB error showed that the error rate levelled off after *ntree*=20, and reached its minimum when *ntree*=52. For the SVM model, various kernel types were assessed, and the linear kernel outperformed both polynomial and RBF kernels. Additionally, the cost (regularization) parameter was tuned for the linear kernel and determined that a value of 1 was the most effective. For GTB, the overall accuracy of GTB classification increased as *ntree* increased until it reached *ntree*=45, then it remained stable and then started to drop. Therefore, the best value for *ntree* in GTB was found to be 45.

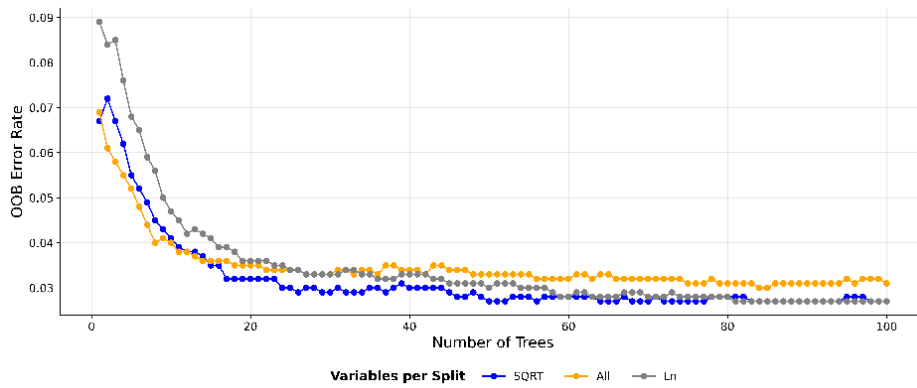


Figure 2.12. Out-of-bag (OOB) error for different numbers of trees and the number of variables (*mtry*) per split was calculated. Different *mtry* per split tested are the square root of the total number of variables (SQRT), the total number of variables (ALL), and the natural logarithm of the total number of variables (Ln).

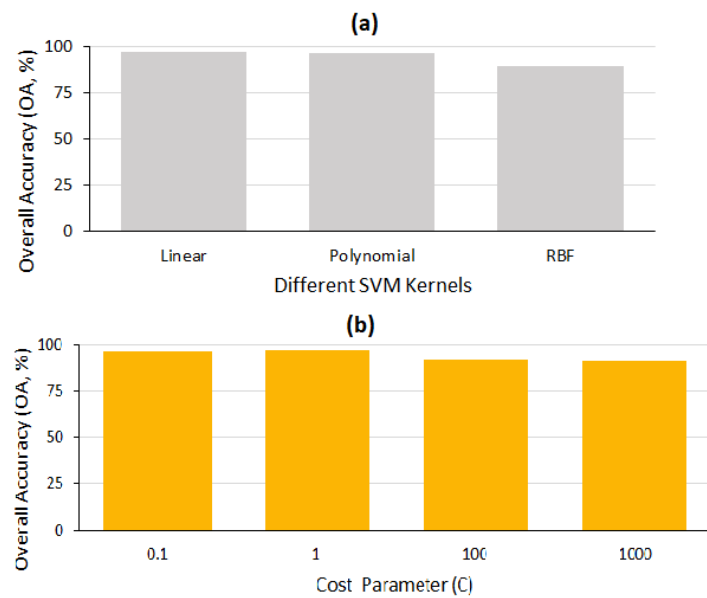


Figure 2.13. (a) Overall accuracy (OA) of classification for different kernel types in SVM model. (b) Grid search to find the best value for the Cost/Regularization parameter for Linear kernel in SVM.

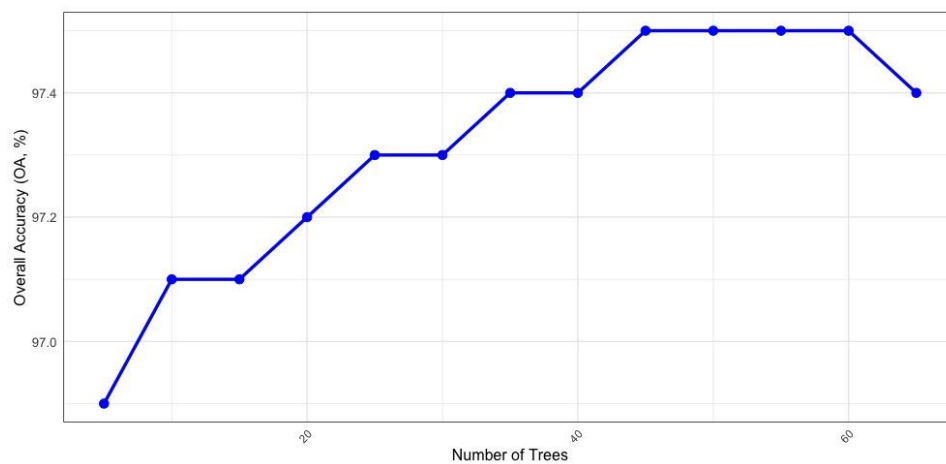


Figure 2.14. The effect of the number of trees (*ntree*) on OA in GBM classification

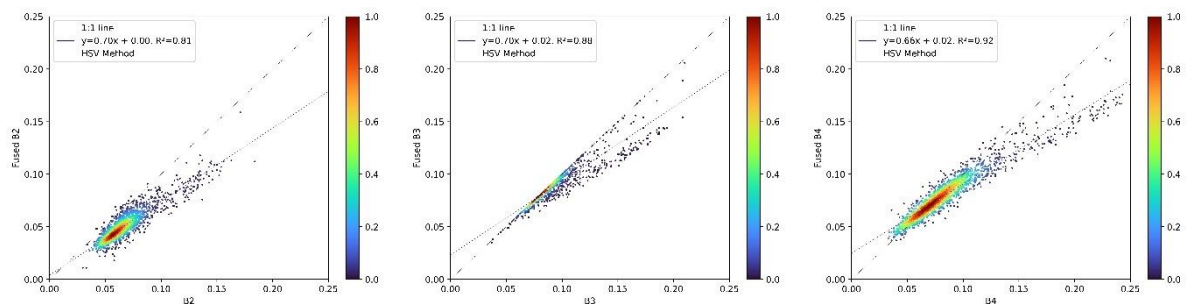
2.4.3. Accuracy Assessment for Sentinel-2+Landsat 8-9 Fused Data

Image fusion was another important evaluation on grasslands classification accuracy, aiming to improve accuracy when an image is temporally, spatially and spectrally enhanced. A case region in the southeastern of Lake Manitoba, which was characterized by a diverse grassland type of tame, native, and mixed samples, was selected to fuse Landsat 8-9 and Sentinel-2 data. The analysis was limited to the period between June 1 and June 16, 2023, for both Sentinel-2 and Landsat 8-9 to minimize potential changes during fusion. The comparison between the Sentinel-2 bands and their corresponding fused bands demonstrated the effectiveness of HSV and BT techniques in preserving the original data's spectral characteristics while increasing the spatial resolution from 30 to 10 m (Figure 2.15).



Figure 2.15. Subsetting areas before and after image fusion for image quality improvement; (a) Landsat 30 m, (b) HSV fused image, (c) BT fused image.

The high r -squared (R^2) values observed across the RGB bands (i.e., B2, B3, B4) in the HSV (0.79, 0.87, and 0.91, respectively) and BT (0.79, 0.62, and 0.94, respectively) indicated that both fusion methods maintained the integrity of the spectral information (Figure 2.16).



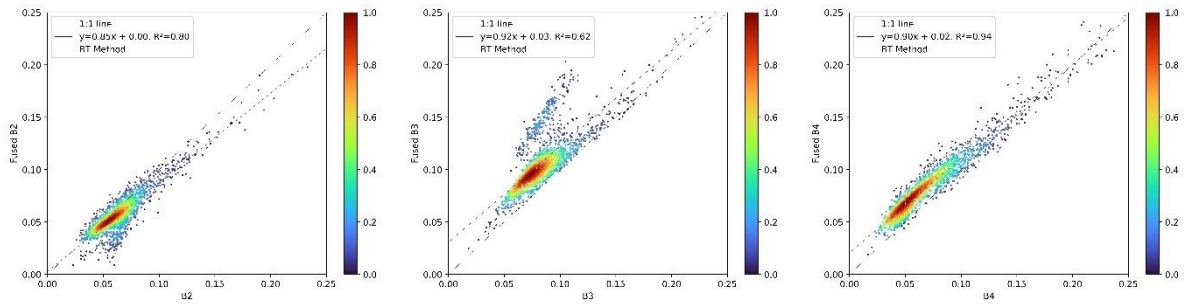


Figure 2.16. Scatter plots of fused and non-fused bands using BT and HSV approaches. Except for band 2 (B2), HSV had a higher R-squared. Around 1900 points were selected to build the scatter plots, and the colour bar represents the point density, ranging from low density (Blue) to high density (Red).

For classification assessment, additional 10 m resolution data, such as NIR band from Sentinel-2 and layers from Sentinel-1, including VV co-polarization and VH cross-polarization, were included in both fused and non-fused images, as shown in Equation (12).

$$\text{Multispectral Image}_{\text{fused}} + \text{NIR}_{\text{Sentinel-2}} + \text{Sentinel-1} \quad (12)$$

Equation (12) shows the input data used by the RF classification model in fused mode. In non-fused mode, only $\text{multispectral Image}_{\text{fused}}$ was replaced with the original MSI (Sentinel-2 in this case), which was then included in the classification process. To classify grassland pixels, non-grassland areas in the selected case region were first screened out using the output of steps 1 and 2 for both fused and non-fused data. Then, all filtered datasets were classified into three grassland types using the RF method, which was shown to deliver the highest performance compared to other machine learning methods. The results indicated that the HSV technique had the highest OA at 59.18%, followed by BT with 58.48%, and MSI with 57.35% (Table 2.7). The Kappa coefficient was 0.36, 0.33 and 0.32 for HSV, BT and MSI approaches, respectively. OA and Kappa values indicated that there was a significant improvement in classification after fusion. PA and PU of mixed class in MSI were the highest compared to HSV and BT. PA of tame grasslands was 77.94%, 73.42% and 60% for HSV, BT and MSI, respectively, and this metric for native class was 61.45%, 68.82% and 72.09%. UA of tame class was 61.63%, 65.17% and 63.93% in HSV, BT and MSI, respectively, while for native class it was 60%, 56.64% and 53.45%. The grassland classifications of fused data and non-fused data are shown in Figure 2.17.

Table 2.7. Comparison of classification performance for fused images using BT and HSV fusion techniques, and non-fused images. The analysis included user's and producer's accuracy of tame, native and mixed classes.

Accuracy Index	Class	MSI	HSV	BT
----------------	-------	-----	-----	----

PA(%)	Tame	60.00	77.94	73.42
	Native	72.09	61.45	68.82
	Mixed	30.19	26.67	17.31
UA(%)	Tame	63.93	61.63	65.17
	Native	53.45	60.00	56.64
	Mixed	59.26	48.00	40.91
OA(%)	-	57.35	59.18	58.48
Kappa	-	0.32	0.36	0.33

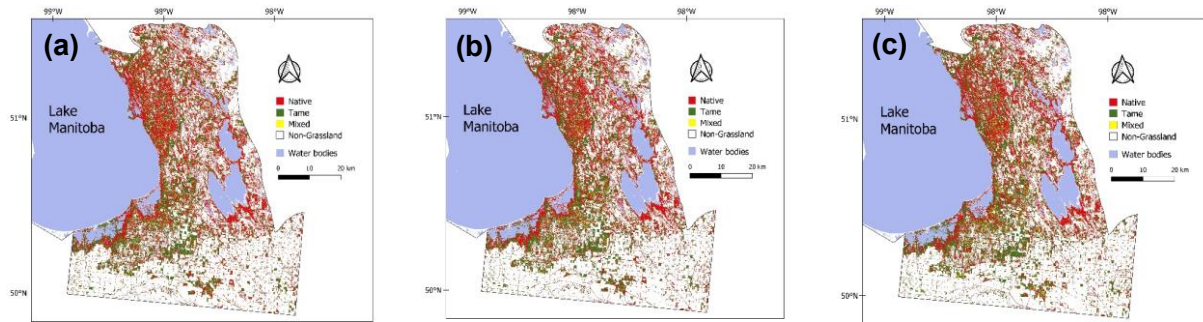


Figure 2.17. The classification maps of pixel-level fusion with RF approach using (a) the multispectral image, (b) the HSV fused image, and (c) the BT fused image.

2.4.4. The Detailed Grassland Map in Manitoba's PE

The Sentinel-1+Sentinel-2 derived grassland map layout using the RF method in Manitoba's PE has demonstrated that the RF model, with an OA of 72.62% and a Kappa coefficient of 0.58, had the highest accuracy compared to the SVM and GTB methods. The Manitoba grassland inventory (MGI) was proposed through GEE, which classified Manitoba's PE into native, tame and mixed classes using the RF method as a better classifier. To produce this map, 1,246 MSI and around 180 SAR images were processed. Considering the computation volume, including preprocessing, cloud masking, aggregating, training, and testing the models, this task would be highly time-consuming for local computational power. However, this huge amount of data was efficiently handled, and a detailed grassland raster dataset was computed in a reasonable time (i.e., less than one hour).

The results of the final step of this study indicated that only 28.55 percent of the total study area consisted of grasslands. The tame class represented the biggest grassland type, accounting for 53.18 % (1,057,436.78 ha) of the total grassland class, followed by the native

class, comprising 25.69 % (510,886.72 ha). The mixed class was the least abundant, with only 21.13 % (420,067.52 ha).

Regarding spatial allocation, the main areas with native grasses were found around Lake Manitoba; some patches were located in the central and western parts of PE (Figure 2.18). The prevalence of tame grasses reflects various land use practices like cultivation or pasture management. The mixed-class areas likely include regions with a combination of native and tame grasslands and other land cover types, such as those interspersed with non-grassland elements like agricultural fields or transitional lands. These areas appear to be scattered across the central and western parts of the PE region, with particular concentration around the west of Lake Manitoba and some isolated patches in the central and southwestern regions. The area around the city of Winnipeg has a tendency towards non-grassland land use.

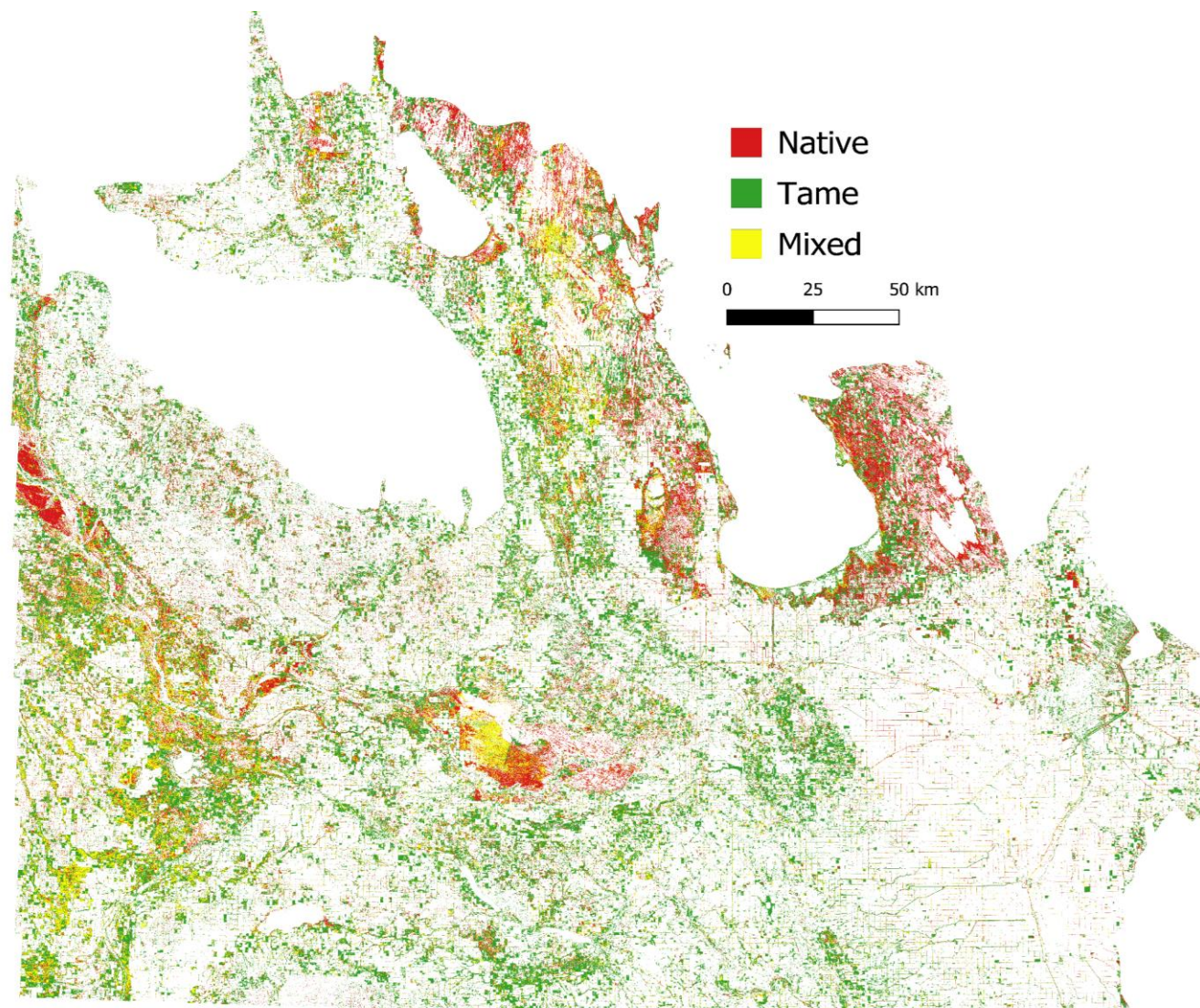


Figure 2.18. The detailed grassland classification of Manitoba's PE using the RF classification machine learning model and Sentinel-1 and Sentinel-2 data combination.

2.4.5. Landscape-Based Grassland Statistics

A breakdown of how much each ecoregion contributed to the total grassland area for each class (native, tamed, mixed) in the three dominant ecoregions in Manitoba's PE is shown in Figure 2.19. Aspen Parkland had the largest grassland area among all ecoregions, with a dominance of tame grasslands (54.14% of the grassland area), followed by mixed grasslands (24.16%) and native grasses (21.70%). The 54.14% of the tamed grassland reflected agricultural activities combined with native ecosystems. In the Lake Manitoba Plain, 50.89% of the grassland area was covered by tamed grasses, demonstrating a higher degree of agricultural modification compared to native systems. The 17.38% of mixed grassland indicated a low interaction between natural and managed ecosystems. Meanwhile, the 31.7% of native grassland suggested a lesser, yet still significant, proportion of natural grassland, pointing to ongoing ecological preservation alongside agricultural use. The Southwest Manitoba Uplands exhibited the lowest overall grassland coverage, with 2.42% of the total grassland. This ecoregion consists of two distinct uplands, including The Pembina Hills and Turtle Mountain sections (R. E. Smith et al., 1998). Pembina Hills Upland is located in south-central Manitoba and presents dominantly a mosaic of farmland, while Turtle Mountain is commonly covered with trees (R. E. Smith et al., 1998).

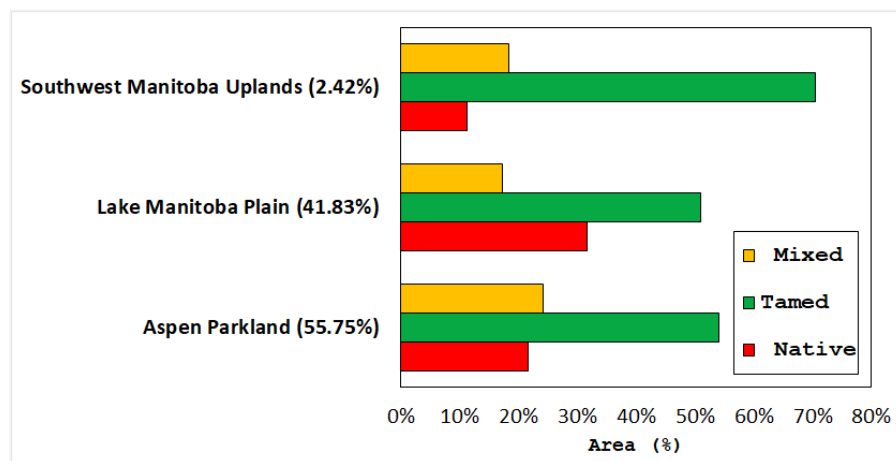


Figure 2.19. Distribution of mixed, tamed, and native grasslands across three ecoregions, highlighting the areas covered by each grassland class. The percentage listed for each ecoregion shows its proportion of the total grassland area, with Southwest Manitoba Uplands at 2.42%, Lake Manitoba Plain at 41.83%, and Aspen Parkland at 55.75%. The relative dominance of each grassland type across Aspen Parkland, Lake Manitoba Plain, and Southwest Manitoba Uplands illustrates regional differences in land use and ecological composition.

In these Uplands, 11.24% of the grasslands were classified as native grasslands. The tame grasslands class, likely located in Pembina Hills, accounted for 70.38% of the grasslands.

The remaining 18.38% comprised mixed grasslands, indicating minimal integration between natural and tame systems.

The newly developed web map is an important geospatial tool for visualizing and analyzing grassland ecosystems in Manitoba. It provides highly reliable spatial data on various grassland types, aiding in understanding their extent, distribution, and characteristics. This supports evidence-based decision-making, allowing the end users to explore data layers, perform spatial analysis, and identify trends related to grassland conservation, land-use planning, and ecosystem services. In addition, the MGI web map is structured in accordance with the FAIR principles (i.e., Findable, Accessible, Interoperable, and Reusable), thereby guaranteeing that the geospatial data it showcases is readily available for a wide spectrum of users and applications (Wilkinson et al., 2016). This fosters collaborative endeavours among governmental agencies, NGOs, and researchers committed to conserving Manitoba's biodiversity and agricultural sustainability, while also addressing issues related to climate change and ecosystem resilience.

2.4.6. Implications to Policy Approaches for Sustainable Grassland Management

Grasslands are widely recognized for their importance in species and habitat conservation, livelihoods for producers and the livestock sector, and cultural importance (Bardgett et al., 2021). They are increasingly understood to be landscape systems that can support national and international biodiversity goals and targets, contribute to regional or national policy objectives and mitigate climate change (Drever et al., 2021). Consequently, a broad range of users and stakeholders, including producers, ecologists, governments and citizens, have a stake in grassland sustainability.

Policy support for grasslands in Canada is largely indirectly applied through federal and/or provincial agricultural, climate or biodiversity policy, incentives, or other user-driven mechanisms (Ayambire et al., 2022). The indirect nature of this guidance can generate uncertainties about whether policy objectives are being met and appropriately reported on. Activities related to grassland decision-making lie primarily within the management of private lands, under greater or lesser influence by public policy.

The use of spatial analysis tools, such as those presented here, can help support a policy framework for grassland monitoring, management and conservation. Greater environmental

and socioeconomic certainty could be realized through: 1) grassland quantification and inventories, i.e., what we have and where it is; 2) change detection and related risk assessment, i.e., how much change is occurring and how quickly; and 3) linking spatial analysis to the policy environment, i.e., what is driving change. Remote sensing can provide the large-scale assessments required to capture large geographical and temporal scales effectively (J. Sun et al., 2022).

2.5. Discussion

There have been limited studies on the classification of large-scale grasslands in Canada using various remote sensing techniques. Hill et al. (Hill et al., 2000) tested the application of RADARSAT to monitor 17,000 ha of native and cultivated rangeland at Onefour, Alberta, Canada. Fisher et al (Fisher et al., 2018) evaluated using LiDAR to classify native and tame grassland types over a 218.15 km² area in southwestern Saskatchewan, Canada. However, this study focused on the extensive region rather than a test study area. This study utilized around 1430 scenes generated by Sentinel-1 and Sentinel-2 from May to October to classify grasslands in Manitoba's PE. GEE allowed us to run the large computations very fast.

Several studies indicated that integrating SAR Sentinel-1 data with optical data like Sentinel-2 enhances accuracy in land use and land cover classification. For instance, researchers (Tavares et al., 2019) demonstrated that the integration of Sentinel-1 and Sentinel-2 for land cover classification achieved a higher accuracy compared to using each sensor individually. Similarly, Borges et al. (Borges et al., 2020) indicated that a combination of multisensory data (Sentinel-1 and Sentinel-2) improved the OA in mapping grasslands and classes like cropland, woodland, water and forest. Our results confirmed that Sentinel-1+Sentinel-2 at both steps one and two classifications attained higher accuracy compared to the OA obtained using either Sentinel-1 or Sentinel-2 individually, as well as in comparison to the Landsat 8-9 and Sentinel-1 combination. For example, in step 1 classification, PA-UA metrics for the vegetation class increased from 94.86%-91.43% with Sentinel-1 or 96.79%-97.37% with Sentinel-2 to 98.59%-97.67% when combined with Sentinel-1 and Sentinel-2. This combination also obtained slightly higher accuracy than the combination of Sentinel-1 and Landsat 8-9 (96-83%-96.84%).

This newly proposed strategy for classifying grassland involves a three-step process to prevent pixels from being inaccurately categorized. For instance, clustering similar vegetation classes together improved accuracy. Similarly, treating all grasslands as a single class yielded comparable results. Initially, attempting to classify all classes in one step led to unsatisfactory results, with several grassland pixels incorrectly assigned to the cropland class. However, adopting the new strategy allowed for a more accurate classification of grassland types. The OA of the grassland class was around 98%, indicating a reliable distinction between grassland and non-grassland classes. Step 1 classification refined the method by reducing noise and removing irrelevant data, such as water and built-up areas.

Mapping grassland types, including tame, native and mixed, is crucial for grassland management (Ren et al., 2008) and for grassland conservation (Yaxing & Quangong, 2001). The accuracy of grassland classification was influenced by the classification machine learning model used. The type of supervised classification machine learning model significantly influenced the accuracy of classifying grassland. During the grassland classification at step 3, three popular machine learning methods were tested. RF performed better than the GTB and SVM methods, before or after ANOVA feature selection. ANOVA helped RF to reduce the number of features from 89 to 61 while simultaneously enhancing the classification accuracy, increasing the OA from 69.96% to 72.62%. In contrast, the GTB method required a more significant number of features, achieving its best OA with 76 features. GTB experienced the greatest improvement in accuracy among all methods, with a 4.49% increase in OA. The accuracy improvement also occurred with SVM development. In terms of the number of selected features and OA enhancement, SVM outperformed GTB and RF, respectively; however, its OA remained significantly lower than both methods. The listing of features according to their scores from higher to lower also showed that optical features (like spectral bands and VIs) outperformed SAR features (like VV and VH) in all classifier models, as shown in Figure 2.A1, in the Appendix of this chapter. Regarding the parameters tuning in machine learning models, this study aimed to evaluate some key parameters for each model and the charts. Based on the charts provided in Figure 2.12-14, it is highly likely that the best parameters for the models were selected. However, we acknowledge that a comprehensive hyperparameter tuning process would be necessary to ensure the models operate at their optimal performance for broader applications.

Frank et al. (2022) showed that the spectral characteristics of different types of grasslands were very similar, making it challenging to distinguish between various grassland classes using only their spectral signatures. Spectral overlapping hinders effective classification and is exacerbated by several external factors such as terrain topography (e.g., hills, depressions, riparian areas), soil properties (e.g., soil organic matter, soil moisture contents), and climate (precipitation and temperature) (Chandola & Vatsavai, 2010). Although those factors can be included in the classification process, spectral overlap caused by the temporal nature of certain classes (like grassland types or cropland) is the major issue, which can be resolved by utilizing multi-temporal techniques such as the one used in the present analyses. Researchers (Aghababaei et al., 2021; Wakulińska & Marcinkowska-Ochtyra, 2020) reported that multi-temporal images can provide data on different stages of vegetation phenology, leading to improvements in vegetation classification.

Gauthier & Wiken (2003) reported that the total grassland area was 14,692 km² of Manitoba's PE (around 21%). This study, using remote sensing data, estimated that 28.55% of the study area was covered by grassland. The Earth Observation Team at AAFC has been producing annual crop-type digital maps (ACI) from 2009 to 2023 with an 85% accuracy at a final spatial resolution of 30 m. The AAFC classification product only distinguishes grasslands into native (pixel value of 110) and tame (pixel value of 122) grasses (AAFC Canada, 2013); ACI showed the area of native grasses was 676,899 ha in 2023. Our study aligns with AAFC estimates, as the AAFC product did not account for the mixed class. There might be some overlap between the mixed class in our classification and the native class in ACI. The study also revealed that the Lake Manitoba ecoregion contains the greatest amount of native grasses, primarily concentrated around Lake Manitoba. On the other hand, the Aspen Parkland ecoregion has the largest proportion of tame grasses, predominantly distributed in the central part of the ecoregion. Additionally, the grassland classes around Lake Manitoba and west of PE are highly intertwined. Finding a border to distinguish these classes is challenging. This may indicate that these areas are at risk of transitioning from native grass types to tame types. This research provides a novel classification strategy that can be adapted to other grassland ecosystems globally. The use of multisource remote sensing data, including Sentinel-1, Sentinel-2, and Landsat 8-9 satellites, combined with advanced supervised machine learning classification techniques, offers a flexible and scalable framework for

grassland classification. However, the regions with distinct phenological patterns, spectral characteristics, or topographic conditions may require adjustments to the input datasets. Additionally, the reliance on field-based ground-truth data for training highlights the importance of local data availability to ensure classification accuracy. The next step to further improve grassland mapping accuracy is to explore the transferability of this methodology to other grassland-dominated areas worldwide by integrating region-specific parameters, such as soil properties (e.g., soil organic matter and soil texture), precipitation, and vegetation types, to develop a globally applicable classification strategy.

2.6. Conclusions

This study demonstrated the effective use of remote sensing technologies, mainly through integrating satellite data in mapping grassland types across Manitoba's PE. It also highlighted the critical role of grassland mapping in assessing landscape changes and preserving the remaining natural grasslands within PE. The study presents a three-step approach to gradually map grasslands, resulting in refined grassland pixels before entering step 3, the final step, where they are classified into tame, native, and mixed classes. Sentinel-1 + Sentinel-2 data as an optimal sensor combination were determined using the RF method among different sensor combinations (Landsat 8-9, Sentinel-1 and Sentinel-2). The PA using Sentinel-1 + Sentinel-2 data ranged from 94.77% to 98.86% (for the non-grassland class), while the UA ranged from 97.41% (for the mosaic grassland class) to 99.14% (for the non-vegetation class). Both PA and UA values for all classes using Sentinel-1 + Sentinel-2 data were higher than those achieved with any other sensor combination.

In all three steps, between RF, SVM and GTB, RF achieved the highest accuracy, with OA and Kappa values of 98.08% and 0.97 for the first two steps, respectively, and 72.62% and 0.58 in step three, respectively. The results showed that using the ANOVA technique at step three would result in higher classification accuracy, with OA increases of 2.66%, 3.96%, and 4.49% for RF, SVM, and GTB, respectively, compared to their initial OA values of 69.96%, 56.58%, and 67.69%. Simultaneously, the feature selection technique reduced the number of features from 89 to 61, 67, and 76 for RF, SVM, and GTB, respectively. This study also assessed the effect of fusion on grassland classification. BT and HSV techniques were used to fuse Landsat 8-9 and Sentinel-2 data. Both methods achieved good results, with all R^2 values between the fused and original images exceeding 0.80, except for Band 3 (B3) in the BT

technique. Grassland pixels of fused and non-fused images were classified using the RF method, and their accuracy performance revealed that HSV had the best performance, with an OA of 59.18% and a Kappa value of 0.36. The BT method ranked second, with an OA of 58.48% and a Kappa value of 0.33. Both fusion methods attained higher classification performance than non-fused images, which had an OA of 57.35% and a Kappa value of 0.32. In all ecoregions in PE, tame grasslands were predominant. The area occupied by tame grasses was two times greater than that of native grasses across the whole ecoregion. Native grasslands might be found more in the Lake Manitoba Plain ecoregion, frequently around Lake Manitoba, than in other places (with 31.73% of total grasslands in this ecoregion). Mapping native grasslands is crucial for understanding and conserving biodiversity in the Canadian Prairie, as this ecosystem supports species at risk and provides essential ecosystem services such as carbon sequestration. In the Aspen Parkland and Southwest Manitoba Uplands (e.g., Pembina Hill Uplands) ecoregions, the area of the mixed grassland class exceeded that of the native grassland class. Specifically, in Southwest Manitoba, the mixed class accounted for 18.38% compared to 11.24% for the native class, while in Aspen Parkland, the mixed class made up 24.16% compared to 21.70% for the native class. This indicates that some areas may have a high potential for transitioning from one grassland type to another.

The MGI v2 web map provides a new platform that can facilitate engaging various stakeholders, including land managers, policymakers, and researchers, by delivering interactive and accessible geospatial information. Implementing the FAIR data model ensures that the grassland classification data are widely usable for diverse applications.

2.7. Appendix

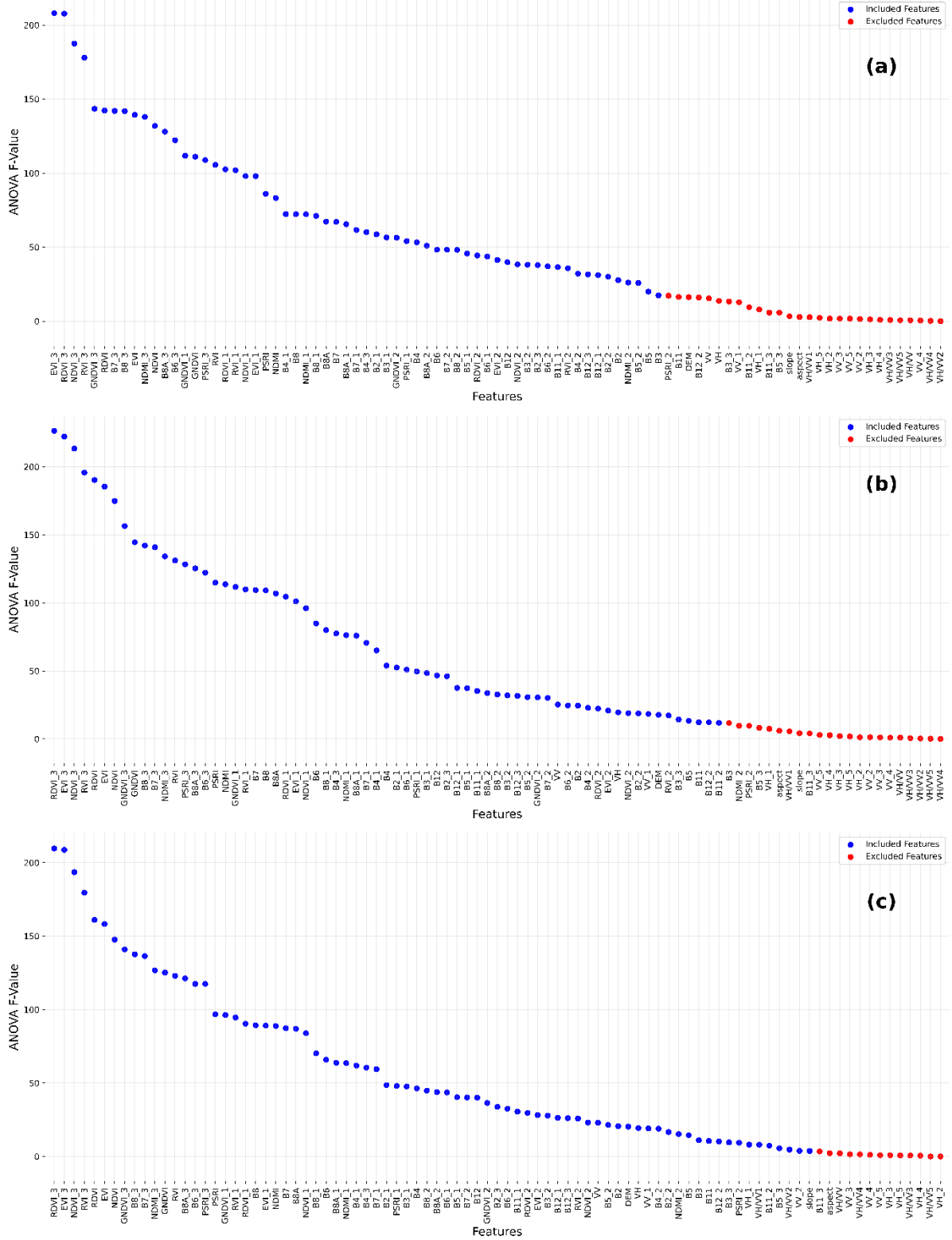


Figure 2.A1. The list of all features with their scores. The numbers following spectral bands, VIs and backscatter variables indicate multiple composite images created during the growing season. Red points show the features that were excluded from classification models to achieve their highest OA and Kappa coefficient; (a) ANOVA F-Value of RF; (b) ANOVA F-Value of SVM; and (c) ANOVA F-Value of GTB.

2.8. References

- AAFC Canada. (2013, November 4). Annual Crop Inventory—Open Government Portal. <https://open.canada.ca/data/en/dataset/ba2645d5-4458-414d-b196-6303ac06c1c9>
- Abdi, A. M. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, 57(1), 1–20. <https://doi.org/10.1080/15481603.2019.1650447>
- Abdollahi, A., Liu, Y., Pradhan, B., Huete, A., Dikshit, A., & Nguyen Tran, N. (2022). Short-time-series grassland mapping using Sentinel-2 imagery and deep learning-based architecture. *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 673–685. <https://doi.org/10.1016/j.ejrs.2022.06.002>
- Abe, S. (2010). Support Vector Machines for Pattern Classification. *Springer*. <https://doi.org/10.1007/978-1-84996-098-4>
- Aghababaei, M., Ebrahimi, A., Naghipour, A. A., Asadi, E., & Verrelst, J. (2021). Vegetation Types Mapping Using Multi-Temporal Landsat Images in the Google Earth Engine Platform. *Remote Sensing*, 13(22), Article 22. <https://doi.org/10.3390/rs13224683>
- Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., & Green, S. (2016). Satellite remote sensing of grasslands: From observation to management. *Journal of Plant Ecology*, 9(6), 649–671. <https://doi.org/10.1093/jpe/rtw005>
- Al-Wassai, F. A., Kalyankar, N. V., & Al-Zuky, A. A. (2011). The IHS Transformations Based Image Fusion (No. arXiv:1107.4396). arXiv. <https://doi.org/10.48550/arXiv.1107.4396>
- Anderson, D., & Cerkowniak, D. (2010). Soil formation in the Canadian Prairie region. *Prairie Soils Crops* 3, 55–64.
- Argenti, F., Lapini, A., Bianchi, T., & Alparone, L. (2013). A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images. *IEEE Geoscience and Remote Sensing Magazine*, 1(3), 6–35. *IEEE Geoscience and Remote Sensing Magazine*. <https://doi.org/10.1109/MGRS.2013.2277512>
- Ayambire, R. A., Pittman, J., Drescher, M., Moreno-Cruz, J., & Olive, A. (2022). Governance of working landscapes: A conceptual framework. *Sustainability Science*, 17(6), 2579–2596. <https://doi.org/10.1007/s11625-022-01178-z>
- Badreldin, N., Prieto, B., & Fisher, R. (2021). Mapping Grasslands in Mixed Grassland Ecoregion of Saskatchewan Using Big Remote Sensing Data and Machine Learning. *Remote Sensing*, 13(24), Article 24. <https://doi.org/10.3390/rs13244972>
- Bardgett, R. D., Bullock, J. M., Lavorel, S., Manning, P., Schaffner, U., Ostle, N., Chomel, M., Durigan, G., L. Fry, E., Johnson, D., Lavallee, J. M., Le Provost, G., Luo, S., Png, K., Sankaran, M., Hou, X., Zhou, H., Ma, L., Ren, W., ... Shi, H. (2021). Combatting global grassland degradation. *Nature Reviews Earth & Environment*, 2(10), 720–735. <https://doi.org/10.1038/s43017-021-00207-2>
- Barrett, B., Nitze, I., Green, S., & Cawkwell, F. (2014). Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using

- machine learning approaches. *Remote Sensing of Environment*, 152, 109–124. <https://doi.org/10.1016/j.rse.2014.05.018>
- Basu, A., & Das, S. (2021). Chapter 22 - Afforestation, revegetation, and regeneration: A case study on Purulia district, West Bengal (India). In G. S. Bhunia, U. Chatterjee, A. Kashyap, & P. K. Shit (Eds.), *Modern Cartography Series* (Vol. 10, pp. 497–524). Academic Press. <https://doi.org/10.1016/B978-0-12-823895-0.00014-2>
- Birth, G. S., & McVey, G. R. (1968). Measuring the Color of Growing Turf with a Reflectance Spectrophotometer. *Agronomy Journal*, 60(6), 640–643. <https://doi.org/10.2134/agronj1968.00021962006000060016x>
- Borges, J., Higginbottom, T. P., Symeonakis, E., & Jones, M. (2020). Sentinel-1 and Sentinel-2 Data for Savannah Land Cover Mapping: Optimising the Combination of Sensors and Seasons. *Remote Sensing*, 12(23), Article 23. <https://doi.org/10.3390/rs12233862>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2(2), 121–167. <https://doi.org/10.1023/A:1009715923555>
- Canada Centre for Remote Sensing. (2010). *Fundamentals of Remote Sensing*. Natural Resources Canada.
- Chandola, V., & Vatsavai, R. R. (2010). *Multi-temporal remote sensing image classification—A multi-view approach*.
- Dao, P. D., Axiotis, A., & He, Y. (2021). Mapping native and invasive grassland species and characterizing topography-driven species dynamics using high spatial resolution hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation*, 104, 102542. <https://doi.org/10.1016/j.jag.2021.102542>
- De Fioravante, P., Luti, T., Cavalli, A., Giuliani, C., Dichicco, P., Marchetti, M., Chirici, G., Congedo, L., & Munafò, M. (2021). Multispectral Sentinel-2 and SAR Sentinel-1 Integration for Automatic Land Cover Classification. *Land*, 10(6), Article 6. <https://doi.org/10.3390/land10060611>
- Drever, C. R., Cook-Patton, S. C., Akhter, F., Badiou, P. H., Chmura, G. L., Davidson, S. J., Desjardins, R. L., Dyk, A., Fargione, J. E., Fellows, M., Filewod, B., Hessing-Lewis, M., Jayasundara, S., Keeton, W. S., Kroeger, T., Lark, T. J., Le, E., Leavitt, S. M., LeClerc, M.-E., ... Kurz, W. A. (2021). Natural climate solutions for Canada. *Science Advances*, 7(23), eabd6034. <https://doi.org/10.1126/sciadv.abd6034>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012a). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>

- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012b). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>
- Earth, G. (2024, January 17). All Clear with Cloud Score+. *Google Earth and Earth Engine*. <https://medium.com/google-earth/all-clear-with-cloud-score-bd6ee2e2235e>
- Encabo, J. B. M., Cordeiro, M. R. C., Badreldin, N., McGeough, E. J., & Walker, D. (2023). Assessment of remotely sensed inventories for land cover classification of public grasslands in Manitoba, Canada. *Grass and Forage Science*, 78(4), 590–601. <https://doi.org/10.1111/gfs.12631>
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., & Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45(2). <https://doi.org/10.1029/2005RG000183>
- Fatholouloumi, S., Firozjaei, M. K., Li, H., & Biswas, A. (2022). Surface biophysical features fusion in remote sensing for improving land crop/cover classification accuracy. *Science of The Total Environment*, 838, 156520. <https://doi.org/10.1016/j.scitotenv.2022.156520>
- Filipponi, F. (2019). Sentinel-1 GRD Preprocessing Workflow. *Proceedings*, 18(1), Article 1. <https://doi.org/10.3390/ECRS-3-06201>
- Fisher, R. J., Sawa, B., & Prieto, B. (2018). A novel technique using LiDAR to identify native-dominated and tame-dominated grasslands in Canada. *Remote Sensing of Environment*, 218, 201–206. <https://doi.org/10.1016/j.rse.2018.10.003>
- Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185–201. [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4)
- Frank, T., Smith, A., Houston, B., Lindsay, E., & Guo, X. (2022). Differentiation of Six Grassland/Forage Types in Three Canadian Ecoregions Based on Spectral Characteristics. *Remote Sensing*, 14(9), Article 9. <https://doi.org/10.3390/rs14092121>
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189–1232.
- Gauthier, D. A., & Wiken, E. B. (2003). Monitoring the Conservation of Grassland Habitats, Prairie Ecozone, Canada. *Environmental Monitoring and Assessment*, 88(1), Article 1. <https://doi.org/10.1023/A:1025585527169>
- Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, 32, 75–89. <https://doi.org/10.1016/j.inffus.2016.03.003>
- Gibson, D. J. (2009). *Grasses and Grassland Ecology*. Oxford University Press.
- Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sensing of Environment*, 58(3), 289–298. [https://doi.org/10.1016/S0034-4257\(96\)00072-7](https://doi.org/10.1016/S0034-4257(96)00072-7)

- Godfree, R., Firn, J., Johnson, S., Knerr, N., Stol, J., & Doerr, V. (2017). Why non-native grasses pose a critical emerging threat to biodiversity conservation, habitat connectivity and agricultural production in multifunctional rural landscapes. *Landscape Ecology*, 32(6), 1219–1242. <https://doi.org/10.1007/s10980-017-0516-9>
- He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. *IEEE Transactions on Knowledge and Data Engineering*. <https://doi.org/10.1109/TKDE.2008.239>
- Henderson, N., & Sauchyn, D. (2008). Climate Change Impacts on Canada's Prairie Provinces: A Summary of Our State of Knowledge. *Prairie Adaptation Research Cooperative*.
- Hill, M. J., Smith, A. M., & Foster, T. C. (2000). Remote Sensing of Grassland with RADARSAT; Case Studies from Australia and Canada. *Canadian Journal of Remote Sensing*, 26(4), 285–296. <https://doi.org/10.1080/07038992.2000.10874779>
- Hubert-Moy, L., Thibault, J., Fabre, E., Rozo, C., Arvor, D., Corpetti, T., & Rapinel, S. (2019). Mapping Grassland Frequency Using Decadal MODIS 250 m Time-Series: Towards a National Inventory of Semi-Natural Grasslands. *Remote Sensing*, 11(24), Article 24. <https://doi.org/10.3390/rs11243041>
- Huete, A., Justice, C., & Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, 49(3), 224–234. [https://doi.org/10.1016/0034-4257\(94\)90018-3](https://doi.org/10.1016/0034-4257(94)90018-3)
- Huo, T., Wang, J., Zhang, Y., Wei, B., Chen, K., Zhuang, M., Liu, N., Zhang, Y., & Liang, J. (2024). Temperate grassland vegetation restoration influenced by grazing exclusion and climate change. *Science of The Total Environment*, 912, 168842. <https://doi.org/10.1016/j.scitotenv.2023.168842>
- IDB - Index: Plant Senescence Reflectance Index. (n.d.). Retrieved August 22, 2024, from <https://www.indexdatabase.de/db/i-single.php?id=69>
- Inoue, Y. (2003). >Synergy of Remote Sensing and Modeling for Estimating Ecophysiological Processes in Plant Production. *Plant Production Science*, 6(1), 3–16. <https://doi.org/10.1626/pps.6.3>
- Kaplan, G. (2018). Sentinel-2 Pan Sharpening—Comparative Analysis. *Proceedings*, 2(7), Article 7. <https://doi.org/10.3390/ecrs-2-05158>
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. Springer New York. <https://doi.org/10.1007/978-1-4614-6849-3>
- Landsat Algorithms | Google Earth Engine. (n.d.). Google for Developers. Retrieved October 26, 2024, from <https://developers.google.com/earth-engine/guides/landsat>
- Liu, Y., Zhang, Z., Tong, L., Khalifa, M., Wang, Q., Gang, C., Wang, Z., Li, J., & Sun, Z. (2019). Assessing the effects of climate variation and human activities on grassland degradation and restoration across the globe. *Ecological Indicators*, 106, 105504. <https://doi.org/10.1016/j.ecolind.2019.105504>

- Loukika, K. N., Keesara, V. R., & Sridhar, V. (2021). Analysis of Land Use and Land Cover Using Machine Learning Algorithms on Google Earth Engine for Munneru River Basin, India. *Sustainability*, 13(24), Article 24. <https://doi.org/10.3390/su132413758>
- Lv, C., Lu, Y., Lu, M., Feng, X., Fan, H., Xu, C., & Xu, L. (2022). A Classification Feature Optimization Method for Remote Sensing Imagery Based on Fisher Score and mRMR. *Applied Sciences*, 12(17), Article 17. <https://doi.org/10.3390/app12178845>
- Lyu, X., Li, X., Dang, D., Wang, K., Zhang, C., Cao, W., & Lou, A. (2024). Systematic review of remote sensing technology for grassland biodiversity monitoring: Current status and challenges. *Global Ecology and Conservation*, 54, e03196. <https://doi.org/10.1016/j.gecco.2024.e03196>
- Lyu, X., Li, X., Gong, J., Wang, H., Dang, D., Dou, H., Li, S., & Liu, S. (2020). Comprehensive Grassland Degradation Monitoring by Remote Sensing in Xilinhot, Inner Mongolia, China. *Sustainability*, 12(9), Article 9. <https://doi.org/10.3390/su12093682>
- M, A., Ahmed, S. A., & N, H. (2023). Land use and land cover classification using machine learning algorithms in google earth engine. *Earth Science Informatics*, 16(4), 3057–3073. <https://doi.org/10.1007/s12145-023-01073-w>
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. (2017). Sen2Cor for Sentinel-2. *Image and Signal Processing for Remote Sensing XXIII*, 10427, 37–48. <https://doi.org/10.1117/12.2278218>
- Mandhare, R. A., Upadhyay, P., & Gupta, S. (2013). PIXEL-LEVEL IMAGE FUSION USING BROVEY TRANSFORME AND WAVELET TRANSFORM. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*.
- Mardian, J., Berg, A., & Daneshfar, B. (2021). Evaluating the temporal accuracy of grassland to cropland change detection using multitemporal image analysis. *Remote Sensing of Environment*, 255, 112292. <https://doi.org/10.1016/j.rse.2021.112292>
- Markham, B., Jenstrom, D., Sauer, B., Pszcolka, S., Dulski, V., Hair, J., McCorkel, J., Kvaran, G., Thome, K., Montanaro, M., Pedelty, J., Anderson, C., Choate, M., Barsi, J., Kaita, E., & Miller, J. (2020). Landsat 9 mission update and status. *Earth Observing Systems XXV*, 11501, 96–102. <https://doi.org/10.1117/12.2569748>
- Masek, J. G., Wulder, M. A., Markham, B., McCorkel, J., Crawford, C. J., Storey, J., & Jenstrom, D. T. (2020). Landsat 9: Empowering open science and applications through continuity. *Remote Sensing of Environment*, 248, 111968. <https://doi.org/10.1016/j.rse.2020.111968>
- McInnes, W. S., Smith, B., & McDermid, G. J. (2015). Discriminating Native and Nonnative Grasses in the Dry Mixedgrass Prairie With MODIS NDVI Time Series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(4), 1395–1403. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. <https://doi.org/10.1109/JSTARS.2015.2416713>
- Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. YU. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening.

- Physiologia Plantarum*, 106(1), 135–141. <https://doi.org/10.1034/j.1399-3054.1999.106119.x>
- Molema, T. R., Tesfamichael, S. G., & Fundisi, E. (2025). Optical and radar remote sensing for burn scar mapping in the grassland biome. *Remote Sensing Applications: Society and Environment*, 38, 101548. <https://doi.org/10.1016/j.rsase.2025.101548>
- Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F. A., Schwarz, L.-M., Ghazaryan, G., Schultz, J., Malenovský, Z., & Dubovyk, O. (2022). Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. *Remote Sensing of Environment*, 282, 113262. <https://doi.org/10.1016/j.rse.2022.113262>
- Nicolau, A. P., Dyson, K., Saah, D., & Clinton, N. (2024). Accuracy Assessment: Quantifying Classification Quality. In J. A. Cardille, M. A. Crowley, D. Saah, & N. E. Clinton (Eds.), *Cloud-Based Remote Sensing with Google Earth Engine: Fundamentals and Applications* (pp. 135–145). Springer International Publishing. https://doi.org/10.1007/978-3-031-26588-4_7
- Paltsyn, M. Yu., Gibbs, J. P., Iegorova, L. V., & Mountrakis, G. (2017). Estimation and Prediction of Grassland Cover in Western Mongolia Using MODIS-Derived Vegetation Indices. *Rangeland Ecology & Management*, 70(6), 723–729. <https://doi.org/10.1016/j.rama.2017.05.005>
- Potin, P., Rosich, B., Miranda, N., & Grimont, P. (2016). Sentinel-1 Mission Status. *Procedia Computer Science*, 100, 1297–1304. <https://doi.org/10.1016/j.procs.2016.09.245>
- Prasad, P., Loveson, V. J., Chandra, P., & Kotha, M. (2022). Evaluation and comparison of the earth observing sensors in land cover/land use studies using machine learning algorithms. *Ecological Informatics*, 68, 101522. <https://doi.org/10.1016/j.ecoinf.2021.101522>
- Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W., & O’Sullivan, J. M. (2022). A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Frontiers in Bioinformatics*, 2. <https://doi.org/10.3389/fbinf.2022.927312>
- Reinermann, S., Asam, S., & Kuenzer, C. (2020). Remote Sensing of Grassland Production and Management—A Review. *Remote Sensing*, 12(12), Article 12. <https://doi.org/10.3390/rs12121949>
- Ren, J. Z., Hu, Z. Z., Zhao, J., Zhang, D. G., Hou, F. J., Lin, H. L., & Mu, X. D. (2008). A grassland classification system and its application in China. *The Rangeland Journal*, 30(2), 199–209. <https://doi.org/10.1071/RJ08002>
- Roch, L., & Jaeger, J. A. G. (2014). Monitoring an ecosystem at risk: What is the degree of grassland fragmentation in the Canadian Prairies? *Environmental Monitoring and Assessment*, 186(4), 2505–2534. <https://doi.org/10.1007/s10661-013-3557-9>
- Roujean, J.-L., & Breon, F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51(3), 375–384. [https://doi.org/10.1016/0034-4257\(94\)00114-3](https://doi.org/10.1016/0034-4257(94)00114-3)

- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974, January 1). *Monitoring vegetation systems in the Great Plains with ERTS*. <https://ntrs.nasa.gov/citations/19740022614>
- Roy, D. P., Wulder, M. A., Loveland, T. R., C.e., W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., ... Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154–172. <https://doi.org/10.1016/j.rse.2014.02.001>
- Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., & Hostert, P. (2022). Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sensing of Environment*, 269, 112795. <https://doi.org/10.1016/j.rse.2021.112795>
- Shaik, A. B., & Srinivasan, S. (2019). A Brief Survey on Random Forest Ensembles in Classification Model. In S. Bhattacharyya, A. E. Hassanien, D. Gupta, A. Khanna, & I. Pan (Eds.), *International Conference on Innovative Computing and Communications* (pp. 253–260). Springer. https://doi.org/10.1007/978-981-13-2354-6_27
- Shorthouse, J. (2010). Ecoregions of Canada's prairie grasslands. *Arthropods of Canadian Grasslands*.
- Singh, S., & Tiwari, K. C. (2021). Exploring the optimal combination of image fusion and classification techniques. *Remote Sensing Applications: Society and Environment*, 24, 100642. <https://doi.org/10.1016/j.rsase.2021.100642>
- Smith, L. N., Byrne, A., Hansen, M. F., Zhang, W., & Smith, M. L. (2019). Weed classification in grasslands using convolutional neural networks. *Applications of Machine Learning*, 11139, 334–344. <https://doi.org/10.1117/12.2530092>
- Smith, R. E., Veldhuis, H., Mills, G. F., Eilers, R. G., Fraser, W. R., & Lelyk, G. W. (1998). *Terrestrial Ecozones, Ecoregions and Ecodistricts of Manitoba*.
- Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2001). Classification and Change Detection Using Landsat TM Data: When and How to Correct Atmospheric Effects? *Remote Sensing of Environment*, 75(2), 230–244. [https://doi.org/10.1016/S0034-4257\(00\)00169-3](https://doi.org/10.1016/S0034-4257(00)00169-3)
- Stehman, S. V., & Foody, G. M. (2019). Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, 231, 111199. <https://doi.org/10.1016/j.rse.2019.05.018>
- Stumpf, F., Schneider, M. K., Keller, A., Mayr, A., Rentschler, T., Meuli, R. G., Schaepman, M., & Liebisch, F. (2020). Spatial monitoring of grassland management using multi-temporal satellite imagery. *Ecological Indicators*, 113, 106201. <https://doi.org/10.1016/j.ecolind.2020.106201>
- Sun, B., Qin, P., Li, C., Gao, Z., Grainger, A., Li, X., Wang, Y., & Yue, W. (2024). Integrating vegetation phenological characteristics and polarization features with object-oriented

- techniques for grassland type identification. *Geo-Spatial Information Science*, 27(3), 794–810. <https://doi.org/10.1080/10095020.2023.2250378>
- Sun, J., Wang, Y., Liu, S., Li, J., Zhou, H., Wu, G., & Haregeweyn, N. (2022). Editorial: Patterns, functions, and processes of alpine grassland ecosystems under global change. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.1048031>
- Tavares, P. A., Beltrão, N. E. S., Guimarães, U. S., & Teodoro, A. C. (2019). Integration of Sentinel-1 and Sentinel-2 for Classification and LULC Mapping in the Urban Area of Belém, Eastern Brazilian Amazon. *Sensors*, 19(5), Article 5. <https://doi.org/10.3390/s19051140>
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghayé, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., ... Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–24. <https://doi.org/10.1016/j.rse.2011.05.028>
- Wakulińska, M., & Marcinkowska-Ochtyra, A. (2020). Multi-Temporal Sentinel-2 Data in Classification of Mountain Vegetation. *Remote Sensing*, 12(17), Article 17. <https://doi.org/10.3390/rs12172696>
- Wang, X., Sun, Z., Lu, S., & Zhang, Z. (2022). Comparison of Phenology Estimated From Monthly Vegetation Indices and Solar-Induced Chlorophyll Fluorescence in China. *Frontiers in Earth Science*, 10. <https://doi.org/10.3389/feart.2022.802763>
- Wang, Y., Ma, L., Wang, Q., Wang, N., Wang, D., Wang, X., Zheng, Q., Hou, X., & Ouyang, G. (2023). A Lightweight and High-Accuracy Deep Learning Method for Grassland Grazing Livestock Detection Using UAV Imagery. *Remote Sensing*, 15(6), Article 6. <https://doi.org/10.3390/rs15061593>
- Wang, Z., Ma, Y., Zhang, Y., & Shang, J. (2022). Review of Remote Sensing Applications in Grassland Monitoring. *Remote Sensing*, 14(12), Article 12. <https://doi.org/10.3390/rs14122903>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Willcock, S., Martínez-López, J., Hooftman, D. A. P., Bagstad, K. J., Balbi, S., Marzo, A., Prato, C., Sciandrello, S., Signorello, G., Voigt, B., Villa, F., Bullock, J. M., & Athanasiadis, I. N. (2018). Machine learning for ecosystem services. *Ecosystem Services*, 33, 165–174. <https://doi.org/10.1016/j.ecoser.2018.04.004>
- Williams, D. L., Goward, S., & Arvidson, T. (2006). Landsat. *Photogrammetric Engineering & Remote Sensing*, 72(10), 1171–1178. <https://doi.org/10.14358/PERS.72.10.1171>
- Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, 80(3), 385–396. [https://doi.org/10.1016/S0034-4257\(01\)00318-2](https://doi.org/10.1016/S0034-4257(01)00318-2)

- Wu, N., Crusiol, L. G. T., Liu, G., Wuyun, D., & Han, G. (2023). Comparing Machine Learning Algorithms for Pixel/Object-Based Classifications of Semi-Arid Grassland in Northern China Using Multisource Medium Resolution Imageries. *Remote Sensing*, 15(3), Article 3. <https://doi.org/10.3390/rs15030750>
- Wu, Y., Du, Y., Liu, X., Wan, X., Yin, B., Hao, Y., & Wang, Y. (2023). Grassland biodiversity response to livestock grazing, productivity, and climate varies across biome components and diversity measurements. *Science of The Total Environment*, 878, 162994. <https://doi.org/10.1016/j.scitotenv.2023.162994>
- Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., ... Cook, B. D. (2022). Fifty years of Landsat science and impacts. *Remote Sensing of Environment*, 280, 113195. <https://doi.org/10.1016/j.rse.2022.113195>
- Xu, D., & Guo, X. (2015). Some Insights on Grassland Health Assessment Based on Remote Sensing. *Sensors*, 15(2), Article 2. <https://doi.org/10.3390/s150203070>
- Yan, X., Li, J., Shao, Y., Hu, Z., Yang, Z., Yin, S., & Cui, Iv yuan. (2020). Driving forces of grassland vegetation changes in Chen Barag Banner, Inner Mongolia. *GIScience & Remote Sensing*, 57(6), 753–769. <https://doi.org/10.1080/15481603.2020.1794395>
- Yaxing, W., & Quangong, C. (2001). Grassland classification and evaluation of grazing capacity in Naqu Prefecture, Tibet Autonomous Region, China. *New Zealand Journal of Agricultural Research*, 44(4), 253–258. <https://doi.org/10.1080/00288233.2001.9513482>
- Young, N. E., Anderson, R. S., Chignell, S. M., Vorster, A. G., Lawrence, R., & Evangelista, P. H. (2017). A survival guide to Landsat preprocessing. *Ecology*, 98(4), 920–932. <https://doi.org/10.1002/ecy.1730>
- Yu, H., Zhu, L., Chen, Y., Yue, Z., & Zhu, Y. (2024a). Improving grassland classification accuracy using optimal spectral-phenological-topographic features in combination with machine learning algorithm. *Ecological Indicators*, 158, 111392. <https://doi.org/10.1016/j.ecolind.2023.111392>
- Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., & Arino, O. (2022). *ESA WorldCover 10 m 2021 v200* [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.7254221>
- Zhang, R., Liang, T., Guo, J., Xie, H., Feng, Q., & Aimaiti, Y. (2018). Grassland dynamics in response to climate change and human activities in Xinjiang from 2000 to 2014. *Scientific Reports*, 8(1), Article 1. <https://doi.org/10.1038/s41598-018-21089-3>
- Zhang, T., Bi, Y., Du, J., Zhu, X., & Gao, X. (2022). Classification of desert grassland species based on a local-global feature enhancement network and UAV hyperspectral remote sensing. *Ecological Informatics*, 72, 101852. <https://doi.org/10.1016/j.ecoinf.2022.101852>

- Zhao, Y., Zhu, W., Wei, P., Fang, P., Zhang, X., Yan, N., Liu, W., Zhao, H., & Wu, Q. (2022). Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period. *Ecological Indicators*, 135, 108529. <https://doi.org/10.1016/j.ecolind.2021.108529>
- Zhou, J., Zhang, R., Guo, J., Dai, J., Zhang, J., Zhang, L., & Miao, Y. (2024). Estimation of aboveground biomass of senescence grassland in China's arid region using multi-source data. *Science of The Total Environment*, 918, 170602. <https://doi.org/10.1016/j.scitotenv.2024.170602>
- Zhou, Z.-H. (2021). *Machine Learning*. Springer Nature.
- Zhu, Z., & Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>

BRIDGE TO CHAPTER 3

Chapter 3 builds on the output of Chapter 2, which devised the best datasets and methods to classify and map grassland distribution in the Manitoba Prairie ecoregion. Based on those maps, this next chapter presents a comprehensive methodology for biomass estimation using both traditional parametric models and advanced machine learning algorithms in the Canadian Prairie region of Manitoba. Three remote sensing sensors with different spatial resolutions were employed in this study, and the influence of environmental variables on AGB was also assessed.

3. GRASSLANDS ABOVEGROUND BIOMASS ESTIMATION USING REMOTE SENSING DATA IN MANITOBA'S PRAIRIE REGION

3.1. Abstract

Grasslands, covering approximately 40% of Earth's terrestrial surface, play a pivotal role in both ecological balance and agricultural economies. Aboveground biomass (AGB) is an important indicator of grassland state and productivity. Monitoring grassland AGB using remote sensing data provides extensive, continuous and reliable information for decision-makers. However, cloud cover often obstructs the optical remote sensing observations' ability to acquire effective and continuous vegetation data over the study area. Given the dynamic nature of grassland ecosystems, users need to have access to up-to-date information for making informed economic and management decisions. To address these gaps, Planet developed Biomass Proxy products offering analysis-ready AGB estimates at 10 meters resolution. This study aimed to develop a multi-stage methodology to (1) evaluate the performance of univariate and multivariate regression models; (2) investigate the usefulness of synthetic biomass samples derived from the Biomass Proxy product to augment field data for machine learning models; and (3) assess the influence of environmental factors on grassland AGB estimation in the Prairie regions. The methodological approach used in the analysis was divided into two stages.

In stage one, univariate linear and exponential regression models were fitted by using six vegetation indices (VIs) derived from Sentinel-2 and the Biomass Proxy as input predictors. For the Biomass Proxy products, the optimal temporal offset was found to be seven days prior to the harvest dates. In contrast, due to cloud cover and the coarser revisit time of the Sentinel-2 sensor, the closest available image dates varied from 1 to 11 days before harvest. Among the VI-based models, the NDVI exponential model achieved the best performance ($R^2=0.53$), while the Biomass Proxy-based exponential model achieved even better performance ($R^2=0.61$).

In stage two, the high-temporal Biomass Proxy was used to synthesize biomass samples, enabling the extension of biomass sampling across the growing season of several years. Environmental variables, like precipitation, temperature, and elevation, were included to enhance model robustness. Multiple datasets (PlanetScope at 3 m, Sentinel-2 at 10 m, and

Landsat 8 and 9 at 30 m) were used in conjunction with parametric and advanced machine learning regression techniques, including random forest (RF), support vector regression (SVR) and artificial neural network (ANN). Results indicated machine learning methods outperformed parametric models ($R^2 \approx 0.84$ vs. ≈ 0.64). SVR and RF consistently provided the highest accuracy across three sensors, with maximum R^2 values up to 0.90. At this stage, principal component analysis (PCA) was used to analyze collinear predictors. In linear regression, where correlation-based feature selection was employed to address multicollinearity, PCA produced similar results ($R^2 \approx 0.61$). For machine learning, the RF model with principal components performed similarly to the RF model with all correlated predictors ($R^2 \approx 0.88$), indicating the effectiveness of RF in handling collinear features. Also, k-fold cross-validation gave more precise, generalizable performance estimations in machine learning models. Feature importance analysis indicated that VIs, climatic variables, and elevation contributed approximately 75%, 23%, and 2%, respectively, to AGB estimation, showing the importance and effectiveness of environmental features.

3.2. Introduction

The grassland ecosystem is one of the largest ecosystems in the world (Zeng et al., 2021). Grasslands offer several important ecological functions, such as carbon stock and soil moisture holding capacity (Xu & Guo, 2015). They also contribute to the economy by supplying feed for livestock. Moreover, they support biodiversity and provide habitats for wildlife (Peeters, 2009). Over the past several decades, grasslands have experienced significant degradation and decline due to human activities, such as overgrazing, recreational use, land use dynamics, and soil erosion, which have intensified natural hazards like droughts, wildfires, and brush encroachment (Li et al., 2022; Xu & Guo, 2015). The agent and severity of grasslands breakdown vary according to the geographical location and the types of grasses present (Zhang et al., 2025). In the face of increasing stress factors, such as climatic variability and anthropogenic pressures, accurate and regular monitoring of grassland conditions is essential for their sustainable and adaptive management (Wang et al., 2019).

Aboveground biomass (AGB) is a key biophysical metric and an important indicator of grassland health and management, which is vital for understanding grassland productivity (Wang et al., 2019; Zeng et al., 2021). Quantifying AGB and monitoring its variations are critical for understanding carbon dynamics in grasslands and preserving pasture resources

(Zeng et al., 2021). AGB, typically, is quantified as the aboveground portion of grassland in g/m² or ton/ha (Zeng et al. 2021). In-situ and remote sensing methods represent the two primary approaches used for estimating AGB. In-situ methodologies can be categorized into destructive and non-destructive techniques (Bazzo et al., 2023). Destructive methods for assessing AGB in grasslands involve direct measurement, which requires the harvesting of AGB, followed by subsequent drying and weighing in laboratory settings. While these terrestrial methods provide high estimative accuracy, their practical application is constrained to limited geographic areas due to the substantial time and labour requirements inherent in these processes (Liu et al., 2023). Non-destructive in-situ measurements have been developed for decades, and their primary metrics relate directly to biophysical variables (e.g., plant morphology and spectral properties) of grasslands to AGB. The rising plate meter (RPM) is the most common tool for AGB estimation. An RPM consists of a weighted circular metallic plate and a meter stick through its centre. To measure biomass in a pasture, the plate is placed at the top of the forage canopy, compressing the leaves and stems until they support its weight. Dry matter (DM) is averaged each time a measurement is taken. RPM has limitations when the sward is tall or uneven, and it does not capture within-paddock spatial variability of biomass production, providing only an average estimate (Figure 3.1).



Figure 3.1. The rising plate meter (RPM) is a common tool for grassland AGB estimation (Rankin & Komainda, 2023).

As an alternative to in-situ methods, remote sensing methods have increasingly been used to estimate AGB in various grassland ecosystems over recent decades, which has led to significant advancements in enabling scholars and decision-makers to monitor grasslands continuously and rapidly at different spatial and temporal resolutions (Vawda et al., 2024; Wang et al., 2019). Multispectral and hyperspectral imaging are the two predominant types of optical remote sensing that are used to estimate grassland AGB (Wang et al., 2022). Landsat

series (e.g., MSS, ETM+ and OLI -1 and 2) (Reinermann et al., 2020; Schwieder et al., 2022), Moderate Resolution Imaging Spectroradiometer (MODIS) satellites (e.g., Terra and Aqua sensors) from NASA (Hubert-Moy et al., 2019; Paltsyn et al., 2017), and the Sentinel Constellation of Sentinel-2 from ESA (Abdollahi et al., 2022; Badreldin et al., 2021) are the most frequently used multispectral spaceborne sensors in the field of landscape detection and prediction (Wang et al., 2022). This widespread popularity can be attributed to availability of free data, extensive global coverage, and compatibility with various remote sensing software packages, including SNAP and QGIS. Compared to well-established multispectral platforms, high-quality hyperspectral satellite sensors with comparable spatial resolution have recently emerged (Zhao et al., 2025). The PRecursoRE IperSpettrale della Missione Applicativa (PRISMA), with 239 spectral bands, and the Environmental Mapping and Analysis Programme (EnMap), with 230 spectral bands are two advanced hyperspectral sensors launched by the Italian and German Space Agency, respectively, each featuring a spatial resolution of 30 m (Zhao et al., 2025).

In addition to spectral reflectance obtained through optical remote sensing, specifically in the Green, Blue, Red, and near-infrared (NIR) bands, a variety of vegetation indices (VIs) are commonly used to estimate AGB. Well-known VIs include the Normalized Difference Vegetation Index (NDVI), Difference Vegetation Index (DVI), Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), and the Ratio Vegetation Index (RVI) (Garrouette et al., 2016; Tian et al., 2023). The most widely used indices for biomass estimation have been NDVI and EVI (Dos Reis et al., 2020; Guerini Filho et al., 2020; Paltsyn et al., 2017). Due to their effectiveness in characterizing vegetation states, NASA has generated global NDVI and EVI products (e.g. MOD13Q1, MOD13A1, MOD13A2, MOD13C1, MYD13Q1) with moderate to high temporal resolution (daily, 8 days and 16 days) but relatively low spatial resolution (250 m, 500 m and 1km), which have been widely used by scientists to study global vegetation changes (MODIS Web, n.d.). Liu et al. (2017) analyzed the AGB dynamics trend from 2000 to 2012 using the relationship between MODIS NDVI and actual grassland AGB. Other research has found a notable example of regional-scale AGB estimation using MODIS NDVI. This study, conducted by Gao et al. (2013), estimated grassland AGB across Inner Mongolia from 2001 to 2011 by combining field-measured samples with MODIS-NDVI time series data. To address

spatial heterogeneity and improve estimation accuracy across different grassland types, they developed region-specific statistical models.

Besides MODIS imagery, other multispectral satellite datasets, including Sentinel and Landsat, provide enhanced spatial resolution and a comprehensive selection of spectral bands that are well-suited for grassland AGB in grassland ecosystems (Wang et al., 2022). Li et al. (2021) estimated biomass in Shengjin Lake Wetland, China, using Sentinel-2 data and ensemble algorithms; By using spectral bands, VIs, and texture features, the extreme gradient boosting (XGBoost) method yielded superior results with a root mean square error (RMSE) of 112.425 g/m² and R² of 0.86, compared to RF with an RMSE of 126.571 g/m² and R² of 0.844. Another study in the Pampa Biome in Brazil assessed degradation of this biome due to exotic species and monocultures, which increases the risk of native species extinction (Guerini Filho et al., 2020). The state of the natural grassland biomass was examined using Sentinel-2 imagery across two rotational grazing systems. Two rest intervals of 375 and 750 degree days (DD) were established. Biomass estimation results were satisfactory, with R²=0.51 for 375 DD and R²=0.65 for 750 DD (Guerini Filho et al., 2020).

Cloud cover can often obstruct optical remote sensing observations' ability to acquire effective and timely data over the study area to monitor grasslands (Dusseux et al., 2014). Synthetic aperture radar (SAR), which is an active remote sensing technology, has the ability to penetrate cloud cover and operate independently of weather conditions. Different SAR sensors have been developed over the past decades, which operate at various bands, such as RADARSAT-2 and Sentinel-1 at C band (Staples, 2005; Torres et al., 2012), ALOS-2 at L band (Rosenqvist et al., 2014), and TerraSAR-X and COSMO-SkyMed at X band (Krieger et al., 2007). Some researchers aimed to leverage SAR data for estimating grassland AGB, either by using them alone or by fusing them with optical data. Braun et al. (2018) utilized three types of SAR satellite data, including ENVISAT, ALOS and SSM/I, which represent surface brightness temperature, to build a correlation between radar backscatter intensities and field data for estimating biomass in the low-biomass Savanna ecosystem. They developed different regression models based on different sensor combinations and years, resulting in R² ranging from 0.13 to 0.56 and RMSE between 1.5 and 2.5 tons per hectare. Schmidt et al. (2016) used both Landsat optical imagery and TerraSAR-X (TSX) X-band radar data to develop a statistical model for estimating pasture biomass in northeastern Australia. The model achieved an

adjusted R^2 value of 0.81 in the dry season and 0.74 in the wet season, with mean standard errors of 332 kg/ha and 240 kg/ha, respectively.

Given the dynamic nature of grassland ecosystems, it is essential for various stakeholders, such as ranchers, ecologists and agronomists, to have access to a continuous monitoring system for making informed economic and management decisions (Burger et al., 2024). To address these gaps, Planet's Biomass Proxy offers an innovative method for integrating SAR Sentinel-1 and optical Sentinel-2 data to build an analysis-ready product that indicates AGB dynamics at 10 m spatial resolution (Burger et al., 2024). This dataset is generated using the Cross-Ratio (CR), a radar-based vegetation index, and NDVI, an optical vegetation index, allowing for daily vegetation monitoring regardless of cloud cover. The product offers a relative biomass measure, with pixel values ranging from 0 (absence of biomass) to 1 (i.e., maximum biomass) (Burger et al., 2024). Guillevic et al. (2024) evaluated the effectiveness of Biomass Proxy in quantifying crop biomass dynamics and yield. A correlation was established between Biomass Proxy measurements and the biomass of corn, winter wheat, and soybean fields. The findings confirmed that Biomass Proxy can effectively detect rapid fluctuations in plant growth induced by agricultural practices or environmental stressors, such as nitrogen deficiency and water stress. Notably, fresh biomass measurements of corn and soybean exhibited the highest correlations with Biomass Proxy, with R^2 values exceeding 0.9 and approximately 0.7, respectively.

There are also satellite sensors with higher spatial resolution (<10 m), which are typically commercial satellites that must be purchased by users. High-resolution optical sensors use cameras with smaller Instantaneous Fields of View (IFOV), which leads to reduced swath width and, consequently, lower coverage capacity (Liang & Wang, 2020). This can be a limiting factor when short revisit times are required for monitoring grasslands (Manfreda et al., 2018). Jargalsaikhan et al. (2024) utilized high-resolution PlanetScope imagery (with 3 m spatial resolution) to estimate aboveground dry biomass (ADB) in Mongolian grasslands. The research used linear regression to establish correlations between satellite-derived NDVI and field-measured ADB. Strong correlations were found, resulting in R^2 values from 0.65 to 0.82 for various grassland types, indicating PlanetScope imagery's potential for biomass estimation in diverse grassland ecosystems. High spatial resolution sensors cover smaller areas but provide detailed information. Small-constellations like PlanetScope have revolutionized this

field by offering daily high-resolution imagery with a spatial resolution of 3 m. However, managing and processing such data becomes challenging for regional and global areas (Jargalsaikhan et al., 2025). Some research, such as Jargalsaikhan et al. (2025), aimed to solve this issue by fusing low-resolution sensors like JPSS-VIIRS (375 m spatial resolution) with high-resolution sensors like PlanetScope through spectral harmonization. This approach allows for the integration of the frequent global coverage from low-resolution sensors with the detailed spatial information from high-resolution imagery.

In addition to VIs, which are critical phenological indicators of grass growth conditions, various studies have identified several terrestrial factors that influence AGB in grasslands. These factors include elevation, latitude, longitude, slope, and aspect; meteorological variables, such as precipitation, temperature and solar radiation, as well as soil characteristics, all of which are used as input variables in predictive models (Zhou et al., 2021). Temperature influences the efficiency of nutrients by altering biogeochemical processes (Yang & Piao, 2006; Zhou et al., 2003), precipitation affects soil moisture content, which directly influences the growth of plants (Jobbágy et al., 2002), and topographic factors like elevation, slope, and aspect influence light intensity, moisture distribution, and vegetation patterns, which in turn significantly impact vegetation biomass (Wang et al., 2023). Researchers examined the role of topography and climate, combined with remote sensing data, in estimating grassland AGB in northern China (Yao & Ren, 2024). The findings showed that including topography and climate data improved the R^2 from 0.43 to 0.64, compared to models using only remote sensing data. Similarly, Li et al. (2024) assessed the impacts of climatic factors, anomalies, topographical elements, and soil properties on AGB using 17,421 in-situ ground-truth samples across approximately 4.2 million km² of Chinese grasslands utilizing Sentinel-1 and Sentinel-2, 17 radar and optical features, alongside 14 non-remote sensing variables categorized into soil properties, climatic factors, topographical elements, and climatic factor anomalies (CFA). Mean annual precipitation (MAP) emerged as the most significant predictor, explaining 27% of AGB variation, while EVI from optical data contributed an additional 14%. Elevation accounted for 5% of AGB variation. Among variable groupings, soil properties (soil organic carbon, clay content, and total nitrogen at 0–30 cm), radar variables (VV and VH), and CFA (precipitation, temperature, vapour pressure deficit, and solar

radiation anomalies) exhibited marginal effects, each contributing less than 15% to AGB variation.

The prevalent models utilized for AGB estimation in grassland ecosystems are generally categorized into two frameworks: physical models and statistical models. Physical models are based on establishing a relationship between remotely sensed reflectance data and various vegetation canopy parameters to effectively estimate AGB (Bazzo et al., 2023). Such models, however, simulate the biophysical processes of vegetation (Zeng et al., 2021), adding complexity and requiring numerous vegetation growth parameters, thereby making the process of parameter optimization complex. Therefore, physical models are not applicable for large-scale biomass prediction since this type of modelling requires huge data inputs (i.e., ground truthing field data or in-situ) (Yao & Ren, 2024). Parametric and non-parametric models are a form of statistical model that is based on the regression patterns between various forms of variables (i.e., features and data that contribute to prediction) for prediction. For example, parametric models contain explicit mathematical representations and are further divided into linear and nonlinear models, which are simple and intuitive to target specific spatiotemporal dimensions and highly sensitive to high spatial heterogeneity and diverse grass species (Ali et al., 2016; Ge et al., 2018; Tian et al., 2023; Wang et al., 2022). In comparison, nonparametric models are superior in dealing with complex nonlinear relationships, offering high prediction accuracy, better fitting performance, and good generalization capability with a remarkable advantage of freely learning capability during modeling calibration without relying on pre-assumed functional relationships (Wang et al., 2022).

Commonly used non-parametric models in grassland AGB estimation include random forest (RF), support vector regression (SVR), and artificial neural network (ANN) (Ali et al., 2017; Barrett et al., 2014; Liu et al., 2023; Morais et al., 2021). Li et al. (2024) analyzed six different models (i.e., parametric and non-parametric), including multiple linear regression (MLR), support vector regression (SVR), multilayer perceptron (MLP), elastic net regression (EN), histogram gradient boosting regressor (HGBR), and RF. The analysis revealed substantial differences in predictive performance among the models, with R^2 values ranging from 0.29 for SVR to 0.68 for RF. Similarly, RMSE values varied between 87.92 and 127.50. In a related study, De Rosa et al. (2021) examined a statistical model generalized using a generalized

additive model (GAM) and RF to predict pasture biomass. The accuracy of both models was high, with prediction errors smaller than 20%. The performance of models was quite low when important variables (e.g., soil types and cumulative rain) were not included, and this shows the significance of extensive input features. Overall, the study demonstrated the effectiveness of statistical and machine learning models for accurate pasture biomass estimation at the field scale. Developing models to address the lack of spatially extensive predictions of regional biomass production and biodiversity in grasslands has been a focus of recent research by Muro et al. (2022). They compared feed-forward deep neural network (DNN) and RF approaches to build their regional models to predict biomass and species richness in temperate grasslands. They used features extracted from Sentinel-1 and Sentinel-2, topography variables, and soil types to develop these models. The results demonstrated that Sentinel-1-based predictors performed poorly, and the combination of optical and SAR data did not improve prediction accuracy despite some studies reporting successful predictions using Sentinel-1 data (Naidoo et al., 2019; Wang et al., 2019). The DNN model outperformed RF in biomass estimation, achieving an R^2 of 0.45 using Sentinel-2 surface reflectance. Additionally, the DNN showed superior generalizability, with a higher R^2 in spatial cross-validation (0.38 compared to 0.26 for RF). Defalque et al. (2024) created machine learning models that predicted pasture biomass and dry matter based on the combination of spectral, environmental, and cattle variables. The XGBoost model produced the best accuracy in biomass estimation ($R^2 = 0.76$; RMSE = 1914.8 kg/ha), whereas SVR was most effective in dry matter ($R^2 = 0.77$; RMSE = 822.7 kg/ha). The findings indicate the usefulness of non-linear models and the incorporation of animal attributes in enhancing the monitoring of pasture conditions in grazing systems. Some research, such as Talebizadeh & Moridnejad (2011) and Yuan et al. (2014) endorsed that machine learning methods consistently outperformed traditional statistical approaches.

The lack of comparative studies between different machine learning and parametric models using input vegetation-based variables from different sensors with different spatial resolutions, climatic and elevation factors limits the understanding of how model types and effective variables affect grassland AGB. In order to fill this gap, extensive sampling of biomass at various points in the growing season and among years would be necessary to encompass the temporal and spatial variability of grasslands. As in-situ measurements are time-

consuming and expensive, the effort to find ways to synthetically produce or supplement biomass samples may be a feasible way to facilitate more confident and sustainable AGB modelling. The objectives of this study were (1) to assess grassland AGB productivity using different VIs derived from Sentinel-2, and Biomass Proxy as input remote sensing data and univariate parametric regression models (e.g., linear and exponential methods), (2) to generate synthetic biomass samples from using high-temporal Biomass Proxy product to address limited field samples for machine learning analysis and to better represent biomass variability across the growing season, (3) to comprehensively evaluate grassland AGB using multivariate parametric models (e.g., linear and exponential regression) and machine learning algorithms (e.g., RF, SVR and ANN) by including synthetic data, environmental factors and different VIs derived from different optical sensors, such as PlanetScope, Sentinel-2 and Landsat 8 and 9 (Landsat 8-9), and (4) to assess the role of environment variables (e.g., temperature, precipitation and elevation) in grassland AGB of Manitoba's Prairie Ecozone (PE).

3.3. Materials and Methods

3.3.1. Study Area and Field Biomass Measurements

The study was conducted in Manitoba's PE, the southern part of the province of Manitoba in Canada, as shown in Figure 3.2. The ecozone's natural land cover consists predominantly of diverse grassland landscapes, which have greatly decreased since European settlement in the early 19th century due to extensive agricultural development and expansion (Shorthouse, 2010). An estimated 70% of the natural vegetation in PE, mostly native grasslands, was lost before the 1990s (Riley et al., 2007). A small but significant portion of the PE includes the tallgrass prairie in the Lake Manitoba Plain Ecoregion, which is known as North America's most endangered grassland type (Shorthouse, 2010).

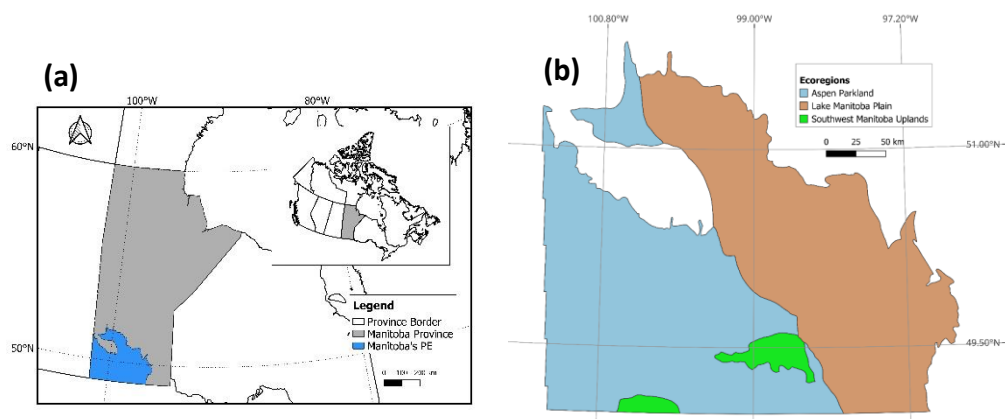


Figure 3.2. Map of the study area in Manitoba, Canada, used for biomass and biodiversity modelling. (a) Location of the Prairie Ecozone (PE) within the province of Manitoba. (b) Distribution of ecoregions within Manitoba's PE, illustrating the spatial variability of ecological zones across the study area.

The PE includes diverse land-use types such as native grasses, tame pastures, haylands, and mixed-use agricultural lands, with agriculture remaining as the predominant land use. The PE is Canada's primary habitat for various grassland species, including blue grama grass, wheat-grasses, alfalfa and sweet clover (Shorthouse, 2010). Although the conversion of natural grassland to agriculture has levelled off over the past few years, different destructive factors such as urban expansion, industrial development, drainage projects, and ongoing agricultural activities threaten landscape integrity through increased fragmentation (Government of Canada, 2002). The study area is characterized by a humid continental climate, featuring warm, moist summers with average temperatures ranging from 15 to 19 °C. The region is predominantly covered by Black Chernozemic soils and has a flat topography (Smith et al., 1998). Its precipitation varies from less than 280 mm/yr in the dry core area to 540 mm/yr in the Lake Manitoba Plain Ecoregion, with a high number of cloud-free days (Smith et al., 1998).

The distribution of biomass samples used in the analyses of the grassland is shown in Figure 3.3. The locations of the fields are spread from northwestern to southeastern Manitoba PE and appear to be distributed across three distinct geographic regions. The number of samples is not evenly distributed in the six fields. The middle field (i.e., NS-L02) has the largest number of biomass samples.

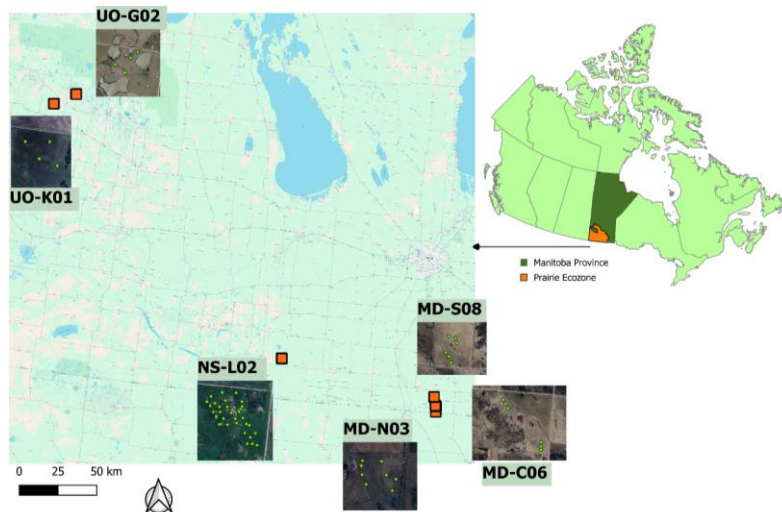


Figure 3.3. Location of six fields where biomass was surveyed in Manitoba's PE (red rectangles) and distribution of biomass samples at each field. Field's name was embedded in them.

Accurate field data is essential to construct a reliable and robust model (Table 3.1). Around 190 AGB samples of grasslands were collected by Agriculture and Agri-Food Canada (AAFC) in the PE of Manitoba using a standardized sampling method. Biomass samples were collected within small plots with a size of 50 x 50 cm. At the "peak" sampling stage, i.e., the biomass of grass reached its maximum, all grasses within a larger area (1.3 x 1.3 m) protected by grazing cages were initially clipped before placing the cages back. Samples harvested were categorized into three types: "Peak" samples were gathered either before grazing occurred or from areas protected by cages; "Regrowth" samples were taken later from areas protected from grazing, containing only new growth without stubble left from earlier clipping; and "Residue" samples, collected only in 2020 without grazing cages, consisted of leftover grass after grazing, combining dead residue with new growth. All sample locations were measured by GPS. Regrowth and Residue biomass samples were discarded due to uneven growth patterns relative to their surrounding areas, resulting in inconsistency between remote sensing features and in-situ measured biomass. Therefore, among biomass samples, 102 peak samples were selected. They were harvested during peak time (early to mid-July) at six fields across three main areas in PE, representing the ecosystem diversity of PE in Manitoba. Further information about sample datasets, including site ID, quarter section, dates and grass types, is provided in Table 3.1.

Table 3.1: The details of grassland sampling sites, such as dates and grass types in PE.

Site ID	Quarter Section	Dates	Grass Types
MD-S08	SE-4-2-4-E1	2021-07-20 2022-07-12	Wet Meadow/Transition; Mixed
	SE-4-2-4-E1	2021-07-20 2022-07-12	Upland Shortgrass; Tame
MD-C06	SE-3-1-4-E1	2021-07-15 2022-07-12	Wet Meadow/Transition; Mixed
	SE-3-1-4-E1	2021-07-15 2022-07-12	Upland Shortgrass; Tame
MD-N03	NE-15-1-4-E1	2022-07-11	Hayland; Tame
	NE-15-1-4-E1	2022-07-11	Upland Tall Grass; Tame
NS-L02-G4	NE-20-3-7-W1	2021-07-13 2022-07-25	Upland Tall Grass-Legume; Tame
	NE-20-3-7-W1	2021-07-13 2022-07-25	Upland Tall Grass-Legume; Tame
	NE-20-3-7-W1	2021-07-13 2022-07-25	Hayland; Tame
UO-G02	NE-8-19-22-W1	2021-07-09 2021-08-01 2022-07-13	Upland Shortgrass; Tame
UO-K01	NW-14-18-24-W1	2021-07-09 2021-08-01	Upland Tall Grass-Legume; Tame

3.3.2. Grassland Biomass Prediction Strategy

Grassland AGB assessment was carried out in three steps. The first was data preparation, acquiring appropriate Sentinel-2 images, Biomass Proxy scenes by considering the temporal difference to in-situ biomass measurements. The selected features were extracted and, alongside Biomass Proxy at different time offsets to the harvesting dates, used as predictors to estimate AGB using univariate parametric models. Step 2 evaluated the predictive capacity of univariate linear and exponential regression models with remote sensing features and assessed the effect of outliers on the models' performance. As field data were limited in both quantity and growing season coverage, the number of samples in various growing seasons and years help to fully capture biomass dynamics. To address this, Biomass Proxy characterized by its high temporal resolution (daily), cloud-free coverage and analysis-ready format was used to generate synthetic biomass samples. Finally, Step 3 built on the outcomes of Step 2 by utilizing the synthetic samples to enable a more comprehensive

evaluation that incorporated climatic and topographic variables known to influence grassland biomass (Li et al., 2024). This step used a broader set of remote sensing sensors (e.g., PlanetScope, Sentinel-2 and Landsat 8-9) and regression models (multivariate linear and exponential parametric methods, as well as RF, SVR and ANN machine learning algorithms). Figure 3.4 illustrates the steps and details about the different sections and components involved in the strategy.

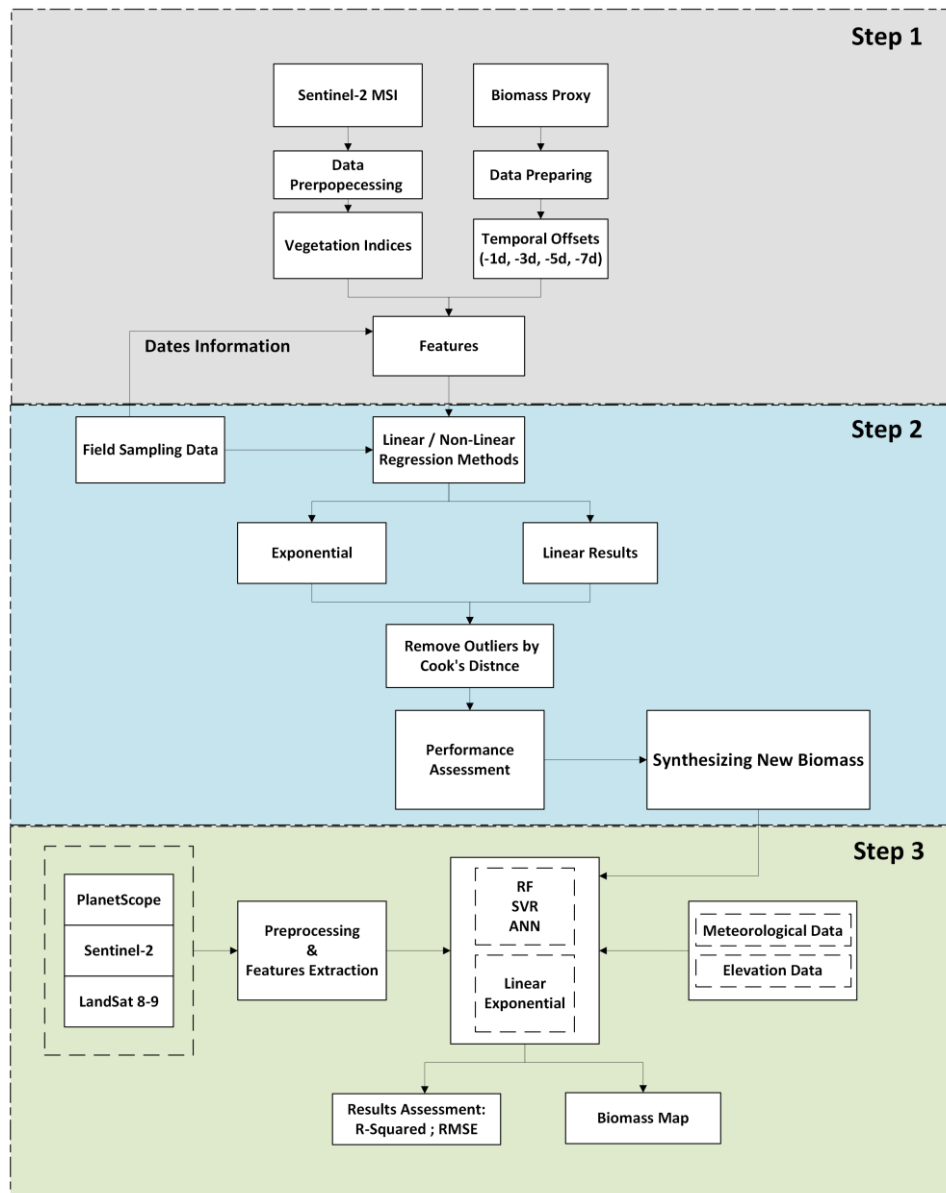


Figure 3.4. General workflow of the primary steps in the grassland AGB estimation strategy. Step 1 was preparing datasets. Vegetation indices (VIs) extracted from Sentinel-2 and Biomass Proxy were acquired according to the timing of ground-truth data collection. In step 2, the fit of several univariate parametric models in AGB prediction was assessed, and Biomass Proxy which offers daily products, was used to generate new biomass samples for a more thorough evaluation on biomass. Step 3 involved a comprehensive evaluation using synthetic biomass samples with multivariate parametric regression models and machine learning algorithms,

incorporating data from three optical satellite sensors (PlanetScope, Sentinel-2, and Landsat 8-9). At this step, the impact of environmental data like precipitation, temperature and elevation was also analyzed.

3.3.3. Optical Remote Sensing Data

Three optical remote sensing sensors, PlanetScope, Sentinel-2 and Landsat imagery with 3, 10, and 30 m spatial resolution, respectively, were used in different parts of this study for grassland AGB estimation. Each sensor offers various levels of image processing, including raw, Top of Atmosphere (TOA), and Bottom of Atmosphere (BOA), also known as Surface Reflectance. This study used the Surface Reflectance image of all the sensors that were atmospherically corrected and provided approximately similar reflectance. To avoid selecting cloudy pixels, a cloud filter was first applied during the downloading of the scenes. Additionally, pixels with NDVI values below 0.2 were filtered out using an NDVI threshold to further reduce the effect of residual clouds or shadows. Even though the revisit times of the sensors were different, attempts were made to pick images that had minimum differences in acquisition dates across sensors to ensure temporal consistency. Sentinel-2 and Landsat images were downloaded from GEE, while PlanetScope images were purchased through their website.

PlanetScope is a constellation of approximately 130 small Earth-observing satellites (known as Doves) operated by Planet Labs. Each satellite is a CubeSat with dimensions of 10×10×10 cm (Frazier & Hemingway, 2021). These satellites provide daily, high-frequency imagery at 3 m spatial resolution, enabling consistent global monitoring (Frazier & Hemingway, 2021). PlanetScope is particularly valuable for applications such as vegetation monitoring, land cover classification, and change detection (Cui et al., 2022; Gašparović et al., 2018; Qayyum et al., 2020). PlanetScope offers several product types: Basic Analytic, Analytic, Visual, and Surface Reflectance, as shown in Figure 3.5. In the study, the data were acquired through a commercial license that was available for the growing seasons from 2021 to 2023.

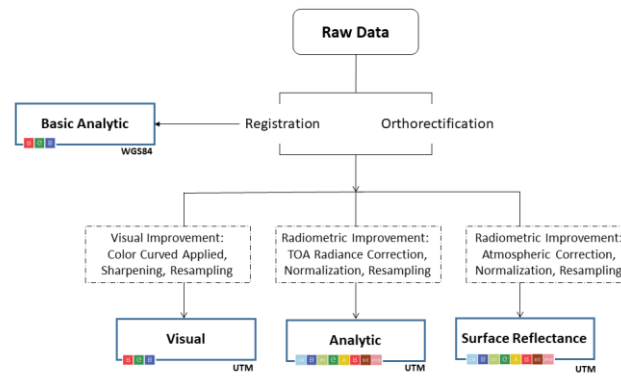


Figure 3.5. Illustrates the processing chain and steps involved to generate each of PlanetScope’s imagery products (PlanetScope | Planet Documentation, 2025).

Sentinel-2 satellite mission consists of a collection of two identical satellites in the same orbit which provide a high revisit frequency of 5 days at the Equator and 10 m spatial resolution (Phiri et al., 2020). Sentinel-2 instrument released two products, which are available to users: Level-1C and Level-2A (Phiri et al., 2020). The atmospherically corrected Level-2A product was used in this study.

Landsat 8-9 obtain seasonal views of global land territory at 30 m resolution for visible, NIR, short wave infrared (SWIR) bands, with an 8-day revisit time when combined (Markham et al., 2020; Roy et al., 2014). Landsat 8-9 data are classified into three Tiers (i.e., categories): Tier 1, Tier 2, and Real-Time data. The highest quality data belong to Tier 1, which meets geometric and radiometric quality requirements, and were used in this study. The comparison of all optical remote sensing data used in this study, in terms of their spectral bands, radiometric and spatial resolution, is provided in Table 3.2.

Surface Reflectance imagery of Sentinel-2 and Landsat 8-9 raster data was acquired from the Google Earth Engine (GEE) catalogue¹, which is a cloud-based platform for Earth science data and analysis that facilitates the analysis and visualization of large-scale geospatial datasets (Parracciani et al., 2024). In GEE, several functions are available to reduce cloud cover, either at the scene level using metadata fields, such as CLOUD_COVER and CLOUDY_PIXEL_PERCENTAGE, or at the pixel level using quality assessment (QA) bands, the cloudScore algorithm and the s2cloudless machine learning model (Harod, 2021; Pasquarella,

¹ <https://developers.google.com/earth-engine/datasets/>

2023). In the case of PlanetScope imagery, the cloud-masking functionality provided on the website of the provider is limited to scene-level filtering by the percentage of overall cloud cover.

Table 3.2. Spectral band characteristics of PlanetScope, Sentinel-2, and Landsat 8–9 OLI sensors used in this study; central wavelength in nanometers (nm), and spatial resolution for each sensor in meters (m), which is the size of the ground area represented by each pixel in the satellite image.

Band Name	Central Wavelength (~nm)	PlanetScope		Sentinel-2		Landsat 8-9 OLI	
		Index Number	Spatial Resolution (m)	Index Number	Spatial Resolution (m)	Index Number	Spatial Resolution (m)
Coastal aerosol	437	B1	3	B1	60	B1	30
Blue	512	B2	3	B2	10	B2	30
Green 1	531	B3	3				
Green	560	B4	3	B3	10	B3	30
Yellow	610	B5	3				
Red	664	B6	3	B4	10	B4	30
Red Edge 1	703	B7	3	B5	20		
Red Edge 2	740			B6	20		
NIR (narrow 1)	781			B7	20		
NIR	833			B8	10	B5	30
NIR (narrow 2)	864	B8	3	B8A	20		
Water vapor	944			B9	60		
SWIR Cirrus	1375			B10	60	B9	30
SWIR 1	1611			B11	20	B6	30
SWIR 2	2191			B12	20	B7	30

Out of approximately 40 VIs utilized in previous studies for the estimation of vegetation biomass due to their sensitivity to the vegetation structure and biomass (Chen et al., 2024; Guerini Filho et al., 2020; Li et al., 2021; Li et al., 2024; Malenovský et al., 2013; Uyeda et al., 2017), six VIs were selected, namely NDVI, EVI, RVI, Renormalized Difference Vegetation Index (RDVI), Green Leaf Algorithm (GLA), and Pan NDVI (PanNDVI). These indices were chosen based on two criteria: (1) their use in prior studies for estimating grassland biomass through remote sensing data and (2) their demonstrated high correlation with in-situ measured biomass samples, as indicated by either Pearson (r) or Spearman (ρ) correlation (Table A1 in the Appendix of this chapter). Six VIs listed in Table 3.3 were calculated based on visible and NIR bands of optical images.

Table 3.3. Vegetation Indices (VIs) were generated using optical remote sensing images and applied in the research models.

Index	Name	Formula
NDVI	Normalized Difference Vegetation Index	$\frac{NIR - R}{NIR + R}$
EVI	Enhanced Vegetation Index	$2.5 \times \frac{(NIR - R)}{(NIR + 6R - 7.5B + 1)}$
RVI	Ratio Vegetation Index	$\frac{NIR}{R}$
RDVI	Renormalized Difference Vegetation Index	$\frac{NIR - R}{\sqrt{NIR - R}}$
GLA	Green Leaf Algorithm	$\frac{2G - R - B}{2G + R + B}$
PanNDVI	Pan NDVI	$\frac{NIR - (R + G + B)}{NIR + (R + G + B)}$

3.3.4. Biomass Proxy Raster Data

One of the newly developed terrestrial variables that helps the understanding of biophysical land cover dynamics is the Biomass Proxy product, which was created based on the combined microwave and optical satellite data to provide a predicted approximation of biomass within high temporal frequency (Burger et al., 2024). The backscatter SAR in VV and VH polarizations (vertical transmit and vertical receive) enables researchers to evaluate three fundamental vegetation characteristics, which include canopy structure, together with vegetation water content and plant tissue dielectric properties (Rubbens et al., 2023; Ulaby et al., 1986). The level of water within vegetation shows a direct connection to biomass; therefore, microwave-based features function as proven indicators of vegetation biomass (Khabbazan et al., 2019; Vreugdenhil et al., 2018). However, radar backscatter is also affected by other factors like soil moisture and incidence angle. To reduce the effects of those factors and enhance radar's capability to detect vegetation, the CR index was introduced (Burger et al., 2024). CR is the ratio between VH and VV extracted from radar Sentinel-1 satellites at 10 m spatial resolution, as shown in Equation (1).

$$CR = \frac{VH}{VV} \quad (1)$$

The optical component depends on Sentinel-2 imagery to derive NDVI for analysis. Chlorophyll concentration detection happens through visible wavelengths yet leaf internal

structural data comes from near-infrared bands, according to Gurney et al. (1993). NDVI is widely used to estimate chlorophyll content in plants, which is directly linked to photosynthetic activity. Chlorophyll absorbs light primarily in the visible spectrum (0.4 to 0.7 μm), while plant cell structures reflect strongly in the near-infrared range (0.7 to 1.1 μm) (Gurney et al., 1993). Thus, optical remote sensing satellites measure both the red and NIR bands to derive NDVI, allowing assessment of vegetation health and density based on the differential reflectance in these two spectral regions. NDVI data is collected at 10 m spatial resolution from the optical Sentinel-2 satellites.

To fuse both radar CR index and optical NDVI in both time and space, it is essential that their values are in the same range. Thus, radar CR is scaled to align with NDVI. NDVI and scaled radar signal (CR) are fused using static weights (fixed contribution) and dynamic weights (based on image quality, cloud cover, and recency). The time series signal is redistributed to daily 10 m resolution biomass maps using spatial weights (Burger et al., 2024).

3.3.5. Landscape Topography Dataset

Topographic data were obtained from NASADEM at a 30 m spatial resolution. NASADEM is a modern version of the Digital Elevation Model (DEM) and associated products generated from the Shuttle Radar Topography Mission (SRTM) data (Crippen et al., 2016). The original SRTM raw signal radar data were reprocessed using improved algorithms and incorporating data primarily derived from the Ice, Cloud, and Land Elevation Satellite (ICESat), Geoscience Laser Altimeter System (GLAS) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) instruments (Crippen et al., 2016). These LiDAR and DEM data were unavailable during the original SRTM processing. The major processing improvements involve void decrease through improved phase unwrapping and using ICESat GLAS data for control (Crippen et al., 2016). The NASADEM products are freely available through the Land Processes Distributed Active Archive Centre (LP DAAC) at one-arc-second spacing (~ 30 m). In this study, NASADEM data were extracted using GEE across biomass sampling fields and resampled to match the spatial resolution of the satellite sensor being used (<https://portal.opentopography.org/datasetMetadata?otCollectionID=OT.032021.4326.2>).

3.3.6. Climatic Data

In this study, Daily Surface Weather Data (Daymet) Version 4 R1, which estimates daily weather parameters on a 1-km grid for North America, Hawaii, and Puerto Rico, was used.

Daymet is derived from selected meteorological station data and various supporting data sources. Each grid has several meteorological variables, including minimum temperature, maximum temperature, precipitation, shortwave radiation, vapour pressure, snow water equivalent, and day length. The dataset covers the period from January 1, 1980, to December 31 (or December 30 in leap years) of the most recent full calendar year for the Continental North America and Hawaii spatial regions. Its spatial resolution is 1 km x 1 km, while its temporal resolution is daily (Thornton et al., 2020).

Daymet Version 4 upgrades the previous versions by implementing multiple enhancements that fix known input weather station bias problems. The key advancements introduce three main modifications to (1) mitigate timing issues in weather station reports, (2) optimize three-dimensional modelling regression algorithms, and (3) establish new temperature bias correction methods in elevated areas (Thornton et al., 2020).

Precipitation and temperature variables were acquired from the Daymet Version 4 R1 dataset through the GEE platform. Some preprocessing steps were applied before using these variables. For the temperature variable, daily average temperature was calculated using daily maximum and minimum air temperature at 2-meter height. The mean of these daily averages over one month prior to biomass sampling dates was computed. For the precipitation variable, daily total precipitation (sum of all forms converted to water-equivalent) was aggregated over the month before biomass sampling dates. The averaging temperature and summing precipitation for a period helped reduce the influences of short-term anomalies. Finally, both climatic variables were resampled before being input into the models.

In Figure 3.6, box plots illustrate the 30-year (1993-2023) monthly averages of temperature and precipitation around the field sites located in UO-G02 and UO-K01 [(a) and (d)], NS-L02 [(b) and (e)], and MD-S08, MD-N02 and MD-C06 [(c) and (f)] farms. The highest variability in precipitation is observed between May and September, corresponding to the main growing season.

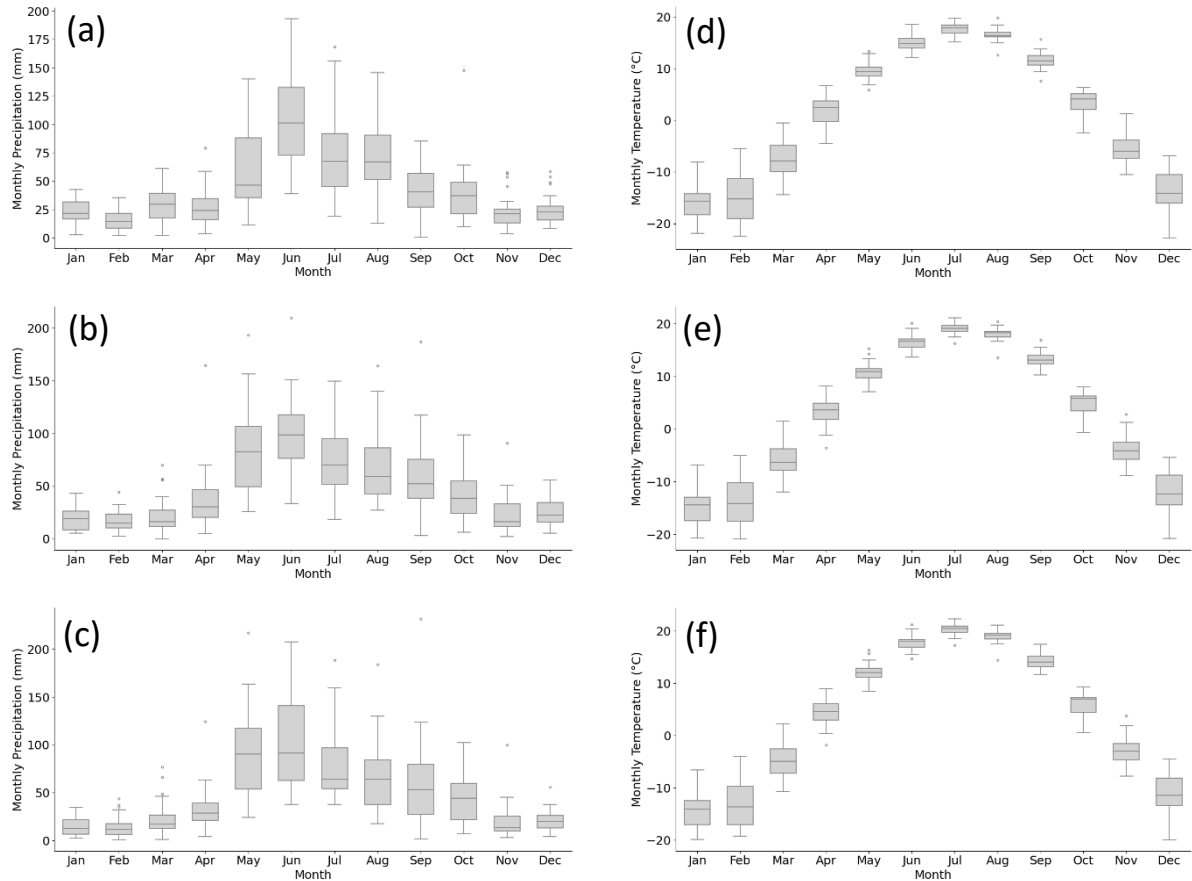


Figure 3.6. Monthly precipitation temperature variation over the 30-year period from 1993 to 2023. Each precipitation box represents the distribution of total precipitation for a specific month across all years, while each temperature box shows the distribution of the monthly average temperature, calculated as the mean of daily maximum and minimum temperatures over the same period. Panels (a) and (d), (b) and (e), and (c) and (f) represent the northern, central, and southern regions of the Prairie regions of Manitoba, respectively.

3.3.7. Parametric Models

In this research, parametric models were developed to investigate the relationship between the AGB of grasslands and predictor features. Two types of parametric regression models, including linear and exponential models, were used to estimate grassland AGB. In the first step, univariate regression models were constructed to build a relationship between 102 AGB samples and remote sensing features. Secondly, after generating synthetic biomass samples (see section 3.3.2), multivariate regression models were used to compare parametric and machine learning approaches for a comprehensive biomass assessment using both environmental variables and VIs. The general forms of the linear and exponential regression models used in this study are shown in Equations (2) and (3), respectively.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2)$$

$$y = \beta_0 e^{\beta_1 x_1} e^{\beta_2 x_2} \dots e^{\beta_n x_n} \varepsilon^n \quad (3)$$

Where y is grassland AGB; the independent variables are $x_1, x_2, \dots, x_i$, $\beta_0, \beta_1, \beta_2, \dots, \beta_i$ represent the parameters of the model; and ϵ refers to the error term.

Using a natural log transformation, it is possible to transform an exponential model to a linear one (Montgomery et al., 2021). The log model, as shown in Equation (4), helps to improve model fit and interpretation.

$$\ln(y) = \ln(\beta_0) + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (4)$$

Ecological modelling extensively uses these models to study different functional relationships between predictor variables and biomass response (Bazzo et al., 2023; Wang et al., 2019; Zhao et al., 2014). They offer several advantages such as simplicity, stability, ease of implementation and interpretability. In addition, they are less exposed to overfitting and require fewer sample data to be trained (Wang et al., 2022). SciPy and Scikit-learn libraries in Python were used to construct parametric regression models.

3.3.7.1. Outlier Detection

In a regression analysis, certain data samples that differ significantly from the majority can lead to incorrect parameter estimates (Sullivan et al., 2021). Extreme observations that cannot be explained by the model are known as outliers (Montgomery et al., 2021). An outlier can occur in either x -space, y -space, or in both dimensions. To gain a deeper understanding of outliers, it is essential to differentiate between two types of observations. The first type is a leverage point, characterized by an unusual X value (i.e., an extreme predictor value), which may significantly affect various properties of the regression model. The second type is an influence point, defined by having both unusual X (predictor) and Y (response) values. An influence point has a significant impact on the coefficients of the regression model and pulls the fitted line toward itself (Montgomery et al., 2021). Leverage of points in a regression model is determined by calculating the hat matrix, as shown in Equation (5). The diagonal elements h_{ii} of the hat matrix are leverage values.

$$H = X(X'X)^{-1}X' \quad (5)$$

Where X refers to the design matrix, containing the values of the independent variables (predictors). To avoid influential data points that have a disproportionate impact on the regression model's parameters, it is important to consider both the location of the point in

the predictor (X) space and its corresponding response (Y) value when measuring influence. Cook (1977) suggested a way to consider both the location of a point (i.e., in the X and Y direction) in measuring its influence. Cook's distance formula is shown in Equation (6):

$$D_i = \frac{(e_i^2)}{p.MSE} \times \frac{h_{ii}}{(1-h_{ii})^2} \quad (6)$$

Where D_i is Cook's Distance for the i -th observation, e_i is the residual for the i -th data point, h_{ii} is the leverage value for the i -th data point, MSE is mean squared error, and p refers to the number of parameters (including the intercept) in the regression model (Bates & Watts, 1988).

As Cook's distance uses both the studentized deleted residual, which is sensitive to Y-outliers, and leverage, which is sensitive to X-outliers, it is sensitive to both types of outliers. There are both benefits and disadvantages to this method. A single statistic allows users to easily pinpoint influential data points, while defining a clear significant cutoff threshold is challenging, and it is not possible to find all X-outliers (Sullivan et al., 2021).

3.3.8. Machine Learning Models

In the last two decades, biomass studies have experienced an increase in the application of machine learning models (Zeng et al., 2019). Three commonly used machine learning methods, including RF, support vector regression (SVR) and ANN, were used to predict biomass. Machine learning methods have shown their high-efficiency capacities for modelling complex, high-dimensional and collinear features (Cutler et al., 2007; Olden et al., 2008). Machine learning algorithms, unlike statistical and empirical models, are non-parametric, which means they do not assume the model to have some fixed form or parameters (Reichstein et al., 2019). This flexibility provides an advantage to them, especially over the traditional statistical methods, when the variables modelled are not independent of each other (Zeng et al., 2019). However, machine learning algorithms suffer from insufficient training due to underrepresented predictor variables and a tendency to converge on local optima, resulting in non-robust or overfitted models (Li et al., 2024; McGovern et al., 2019).

RF, as an ensemble approach composed of a large set of decision trees (DTs) (Breiman, 2001), has the ability to decrease bias and overfitting (Karlson et al., 2015). In the RF algorithm, two parameters should be determined by users. First, the number of trees (*n_{tree}*)

in the regression model must be defined. For each tree, the algorithm constructs a bootstrap sample of the same size as the training set by randomly selecting observations from the training set with replacements. Because of replacements during the selection, some observations appear more than once in each bootstrap sample, resulting in about one-third of the original samples being typically left out. The data included in bootstrap samples are titled “in-bag” data, while excluded observations are known as “Out-of-Bag” (OOB) data (Figure 3.7).

The second key parameter that should be defined in the RF algorithm is the number of features (*mtry*). At each node of the tree, *mtry* variables are randomly selected from the total number of features, and the optimal split is determined based on the lowest Gini Index (GI). All trees are grown using bootstrap samples, with random feature selection, and pruning turned off. To assess model performance, the algorithm computes the mean square error or OOB error by comparing predicted values made using the OOB data and the ‘in bag’ data. The estimate of OOB error provides an unbiased and reliable internal measure of the model’s predictive accuracy, since the OOB data are not used in training or pruning any regression trees in the ensemble. In this study, the model parameters were set to 300 trees (*ntrees*=300) and 9 features per split (*mtry*=9). The 70% of the total observations (approximately 650 labelled points at each satellite sensor) were randomly selected to train the model and the remaining data were used to test the model.

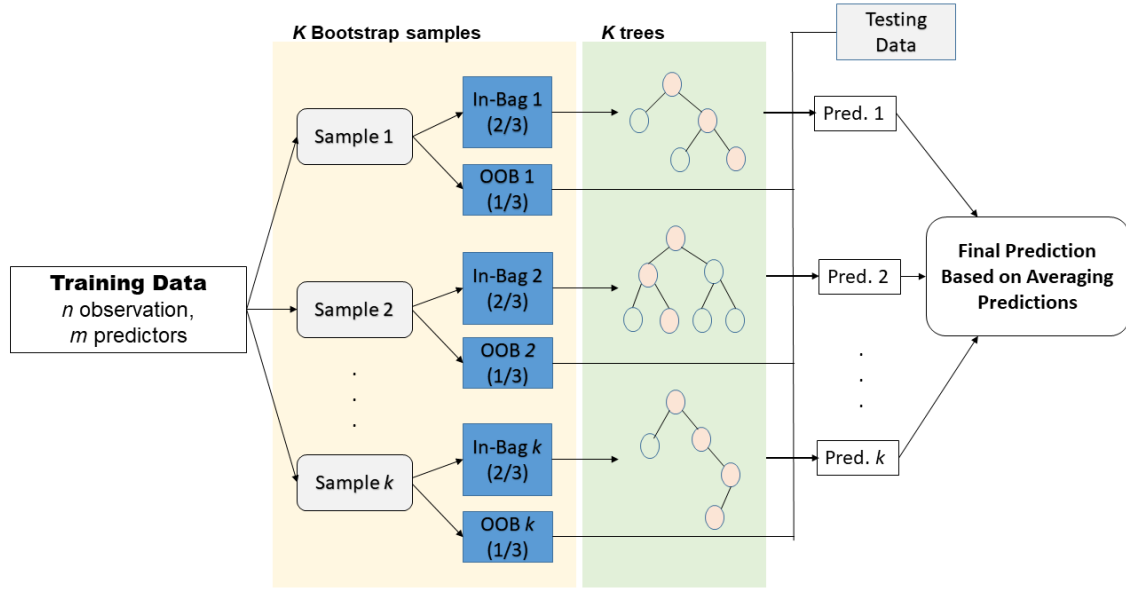


Figure 3.7. Random forest (RF) components in the regression task during the training process. K bootstrap samples are produced using the training data, whereby, for each sample generated, an individual decision tree is built on a random subset of the features. Two-thirds (In-Bag) of each bootstrap sample is used for training and one-third (OOB) is used to test internal validation.

SVR is an extension of SVM. SVR uses the same principle for addressing regression problems (Vapnik, 1999). However, unlike SVM, the idea of hyperplane and margin with a different definition is used in SVR (Rodríguez-Pérez & Bajorath, 2022). Margin in SVR refers to the model error tolerance, also known as the ϵ -insensitive tube, as shown in Figure 3.8. This tube allows slight deviations from the regression hyperplane not to count as errors. The ϵ -insensitive region is the region around the hyperplane where deviations of data points are tolerated and not penalized in the SVR model (Rodríguez-Pérez & Bajorath, 2022). SVR utilizes kernel functions to project input data into high-dimensional feature spaces where linear relationships can be identified and exploited for the regression task (Zocca et al., 2017). The popular Radial Basis Function (RBF) kernel functions were tested in the study (Brereton & Lloyd, 2010). The performance of the SVR model heavily depends on the proper tuning of its hyperparameters, like loss function ϵ and penalty factor C (Wang et al., 2016). In addition, it is important to choose a proper kernel function. This study used the default configuration generated by the related Python library during model implementation. The study used the default configuration with the following parameters: the regularization parameter ($C=100$), the kernel coefficient in the RBF kernel ($\gamma=0.1$), and the margin of tolerance ($\epsilon=0.1$). The 70%

of the total observations (approximately 650 labelled points at each satellite sensor) were randomly selected to train the model and the remaining data were used to test the model.

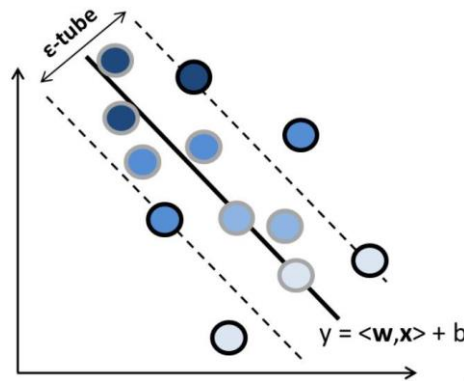


Figure 3.8. The support vector regression (SVR) with ϵ -tube. The central line represents the regression function, while the dashed line defines the ϵ -insensitive margin (ϵ -tube). If data points are placed inside the tube, they are not penalized during training.

The ANN model was constructed using the TensorFlow library within the Python programming environment. This study involved the development of a multi-layer neural network comprising three fully connected hidden layers. The first, second, and third hidden layers comprise 64, 64, and 32 neurons, respectively. The quantity of neurons present in the input and output layers of an artificial neural network regression aligns with the number of input features and the singular target variable, which is AGB in the present analysis. Hidden layers utilize the ReLU activation function [Equation (7)], which is the closest to its biological counterpart (i.e., the manner in which actual neurons within the human brain fire if only the input is bigger than a given threshold) (Haykin, 1994; Zeng et al., 2021). Activation functions allow for modelling non-linear relationships (Figure 3.9). The ReLU function can be described mathematically and visually as follows:

$$ReLU(x) = \max(0, x) \quad (7)$$

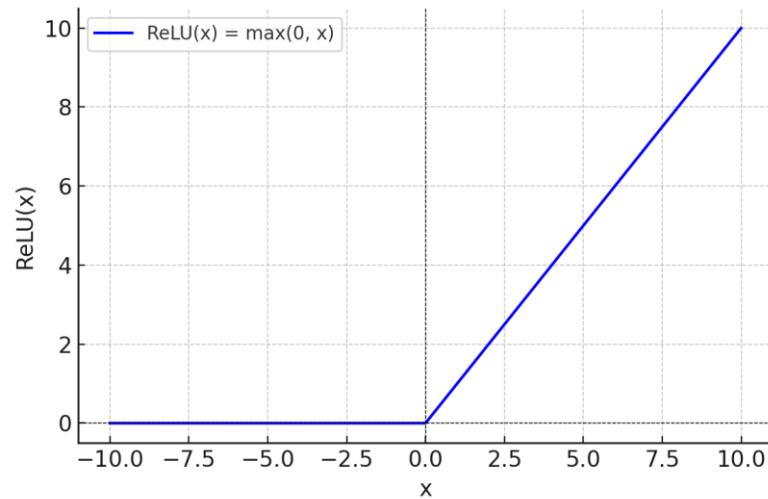


Figure 3.9. The ReLU activation function and its plot. Despite its simple form, ReLU's non-linearity allows it to learn more complex data patterns between features.

The ANN model is trained using the backpropagation step that is handled automatically by TensorFlow (Brownlee, 2016). The study assigned 70% of the total observations for the training part. All input data were normalized using Min-Max scaling to maintain stable input distributions before the training process. The training input data were further split into subsets, including training and validation sets. The training set (70% of training input) contributed to adjusting the weights of neural networks, and the validation set (30% of training input) validated the training process. During training, the error decreases in the validation set before it rises after the network passes its optimal fit point. Training processes generally stop when the validation error finds its minimum point (Bishop, 2007). The testing part (the 30% of the total observation), which was selected randomly, was used to assess the final generalization power of the trained model (Zeng et al., 2021). The number of total observations was approximately 650 labelled points at each satellite sensor that splitting into training and testing sets was performed randomly. The ANN was trained for 100 epochs with a batch size of 32. Model performance during training and validation was monitored through the loss function (MSE).

The regression models (i.e., parametric and machine learning methods) used in this investigation were implemented on the cloud-based Google Colab platform using the Python programming language. A variety of libraries, such as NumPy, pandas, GeoPandas, seaborn, scikit-learn, and TensorFlow, were used in the implementation of the models.

3.3.9. Model Assessment

The study used principal component analysis (PCA), one of the most common dimension-reduction methodologies, to address the issues of correlated predictors in regression models. PCA projected features into a lower-dimensional space, maintaining as much original variance as possible in the process. It attains this by determining new orthogonal axes, known as principal components, which represent the directions of greatest variability in the data (Abdi & Williams, 2010).

Mathematically, the transformation is a process in which the original data matrix X is projected onto a set of orthogonal vectors (eigenvectors) made up of a matrix W . The transformed data matrix Z is then obtained as shown in Equation (8):

$$Z=XW \quad (8)$$

Here, W has the eigenvectors of the covariance matrix of X and each eigenvector represents a principal component. The corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ describe the variance described by each principal component, and they are arranged so that [Equation (9)]:

$$\lambda_1 > \lambda_2 > \dots > \lambda_n \quad (9)$$

After obtaining a smaller subset of the most important principal components, PCA successfully achieves dimensionality reduction while minimizing information loss (Abdi & Williams, 2010).

Cross-validation is a statistical method of evaluating the generalization ability of predictive models, often as an alternative to data lacking in the required quantity. Instead of using the train-test split just once, cross-validation divides the data set into several subsets (also called folds) and trains and tests the model on each one of these subsets in turn. In the most applied form, the so-called k -fold cross-validation, the data is partitioned into k equally sized folds. The model is learned on $k-1$ folds and tested on the remaining one. This is done k times, and each fold is used as the validation set once. The final performance metric (e.g., R^2 , RMSE) can be calculated as the average across all k iterations (Hastie et al., 2009).

Regression performance was evaluated using R^2 or the coefficient of determination and RMSE on the test dataset, which had been split previously from the input data. R^2 measures

how well the model explains the variability of the target variable, while RMSE quantifies the average difference between predicted and observed values [Equation (10-11)].

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (10)$$

$$RMSE = \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n} \quad (11)$$

Where y_i , $\hat{y}_i$ and $\bar{y}_i$ refer to actual in-situ biomass value, predicted biomass value from the model and mean value of in-situ biomass, respectively.

3.4. Results

The histogram of AGB of grassland samples within each biomass range is shown in Figure 3.10. The histogram was also overlaid with a kernel density estimation (KDE) curve. The biomass data show a moderately right-skewed distribution, indicating the presence of more samples with mid-range biomass values and fewer samples with very low or very high values.

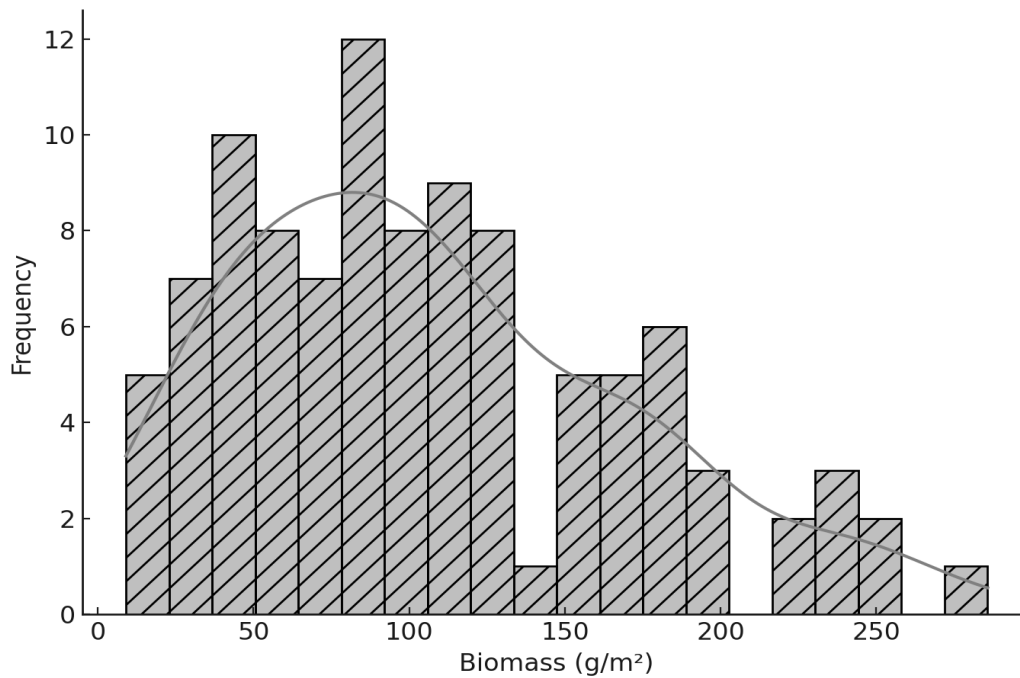


Figure 3.10. Histogram distribution of AGB across validated grassland sampling points, overlaid with a kernel density estimation (KDE) curve.

3.4.1. Grassland AGB Estimation Using In-situ Samples

3.4.1.1. Parametric Models Using Vis as Predictors

Initially, ten Sentinel-2 Level-2A images were selected based on minimal cloud cover and temporal proximity to field biomass sampling, in order to extract the selected VIs as

inputs for the VI-based regression models. A total of 102 grassland AGB samples were collected over six farms on different dates. To ensure temporal consistency with field samples, all images were acquired during the peak growing season (early to mid-July at this research), aligning with field harvests and representing a similar phenological stage. The statistical performance of univariate linear and exponential functions is illustrated in Table 3.4. The results indicated that, for most predictors, the exponential models were found to outperform the linear models, achieving higher R^2 values. In linear models, the RVI predictor showed the best model performance with an R^2 of 0.33, while the EVI predictor had the poorest performance ($R^2=0.20$). There were minimal differences in R^2 among the six indices for biomass estimation using the exponential model. GLA, RVI and PanNDVI explained biomass variance better than the other indices with an R^2 of 0.31, while the lowest value was related to EVI with an R^2 of 0.27.

The Pearson and Spearman correlations are also shown in Table 3.4. Although the Pearson correlation assesses the linear relationships, the Spearman correlations evaluates monotonic relationships (whether linear or not). Linear regression findings were supported by the correlation values, where the highest Pearson correlation coefficient was achieved by RVI (0.58) and the lowest one was achieved by EVI (0.45). Additionally, the results indicated that in the cases where Spearman correlation was higher than Pearson correlation, the relationship between predictors and biomass was likely non-linear, and thus, the monotonic association was stronger than the linear one. For instance, NDVI and RDVI had the largest difference between Pearson and Spearman correlation coefficients, indicating that there was weaker linear correlation with biomass. In line with this, the two indices recorded greater increases in R^2 for exponential models (0.08 and 0.07, respectively) compared to other predictors.

Table 3.4. The coefficient of determination (R^2) of univariate linear and exponential regression models for biomass estimation using the selected vegetation indices (VIs).

Index	R^2 of Linear $y = a + bx$	Pearson correlation (r)	R^2 of Exponential $\ln(y) = a + bx$	Spearman correlation (ρ)
NDVI	0.22	0.47	0.30	0.54
EVI	0.20	0.45	0.27	0.46
RVI	0.33	0.58	0.31	0.54
RDVI	0.23	0.48	0.30	0.50
GLA	0.30	0.54	0.31	0.55

PanNDVI	0.26	0.51	0.31	0.55
---------	------	------	------	------

* The p-value for all indices was less than 0.001.

3.4.1.2. Ground-Truth Outlier Removal

Since the parametric results showed that both linear and exponential models were not sufficiently effective in explaining the variability of the target variable, the poor performance might be attributable either to general model limitations or to a few influential outliers. The possible outliers might be due to either field-side issues, like geolocation mismatch and sampling errors, or remote sensing side issues, like mixed pixels and temporal mismatch. To analyze outliers, Cook's distance was calculated for all pair-observations (i.e., VIs-Biomass pairs) used in the regression models to assess the presence of outliers in the datasets. Although there is no universally strict cutoff threshold for Cook's distance, there are common "rules of thumb" widely used in practice (Sullivan et al., 2021). In this study, pair-observations with $D_i > 4/n$ (i.e., $D_i > 0.04$) were considered influential and removed. This value serves as a good threshold to find highly influential samples as outliers (Lee, 2025). Figure 3.11 shows the Cook's distance values for all observations used in the linear parametric model with RVI predictor. It is important to note that the same outlier removal procedure was applied to all other vegetation index-based regression models. Approximately 10% of the 102 observation pairs were identified as outliers and removed from subsequent analyses.

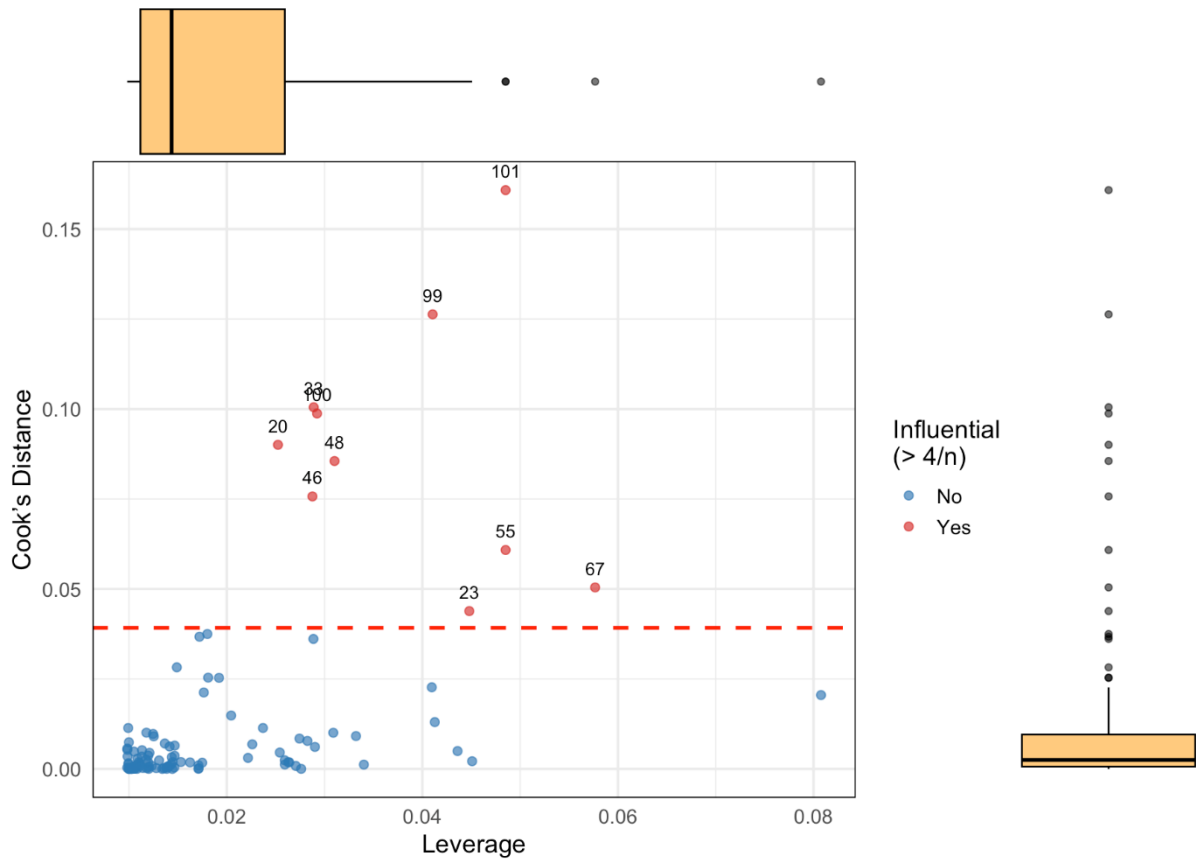


Figure 3.11. Diagnostic plot of leverage and Cook's Distance for the linear regression model using the RVI predictor ($n = 102$). The figure shows the scatter plot of Cook's Distance versus leverage, with the common influence threshold ($4/n$) indicated by a dashed red line. Observations classified as influential (Cook's $D > 4/n$) are highlighted in red and labelled by their observation index. Boxplots of Cook's Distance values are shown along both graph axes.

The results, provided in Table 3.5, show model performance metrics for both situations: with and without outliers. The findings indicated that after removing outliers, similar to the previous step, exponential models generally outperformed linear models. However, an exception was found in RVI and GLA when the linear regression performed better with the $R^2=0.50$ and 0.48 , respectively, compared with the corresponding 0.41 and 0.37 in the exponential models. Among the linear models, the GLA-based regression showed the largest increase in R^2 after outlier removal, with an improvement of 0.18 , followed by RVI with an increase of 0.17 . In the exponential models, when outliers were removed, the exponential NDVI model performed the best, resulting in an R^2 of 0.53 compared to its previous value of 0.30 . The most improved models following NDVI were those based on EVI-, RDVI-, and PanNDVI features, with the increase in R^2 of 0.20 , to the final values of 0.47 , 0.50 , and 0.51 . These results indicated that removing outliers substantially enhanced model performance,

particularly for indices better suited to non-linear relationships with biomass (e.g., NDVI, RDVI, PanNDVI), whereas linear models remained more effective for indices like RVI and GLA

Table 3.5. The comparison between parametric linear and exponential models using vegetation indices (VIs) with/without outlier samples.

Index	Without Outliers		With Outliers	
	R ² of Linear* $y = a + bx$	R ² of Exponential $Ln(y) = a + bx$	R ² of Linear* $y = a + bx$	R ² of Exponential $Ln(y) = a + bx$
NDVI	0.37	0.53	0.22	0.30
EVI	0.34	0.47	0.20	0.27
RVI	0.50	0.41	0.33	0.31
RDVI	0.38	0.50	0.23	0.30
GLA	0.48	0.37	0.30	0.31
PanNDVI	0.40	0.51	0.26	0.31

* The p-value for all indices was less than 0.001.

3.4.1.3. Parametric Models by Biomass Proxy

The capability of the Biomass Proxy was also evaluated using linear and exponential regression models to estimate grassland AGB. Since time-series data of the Biomass Proxy during the growing seasons from 2019 to 2023 were available for all sample points, parametric regression models were developed between biomass measurements and the Biomass Proxy using various temporal offsets from the sampling dates. As shown in Figure 3.12, the linear and exponential models under different scenarios regarding time offset were evaluated, both with and without outlier inclusion.

The findings demonstrated that in linear regression with the Biomass Proxy variable, the model with outliers obtained an increase in R² from 0.18 to 0.26, while the model without outliers showed a greater increase in R² from 0.21 to 0.40 as the offset dates of the Biomass Proxy shifted from 0 to -7 days. Similarly, in the exponential regression model, R² increased from 0.24 to 0.38 with outliers and from 0.46 to 0.61 without outliers over the same temporal shift.

Overall, these results indicate that a seven-day offset before the harvest date was optimal for both regression types, yielding the highest performance in biomass estimation. Meanwhile, when compared to all the other VI-based models, the exponential Biomass Proxy model with a seven-day offset and without outliers had the largest R² value of 0.61 compared to the rest of the linear and non-linear models. This indicates the substantial effectiveness

and reliability of the Biomass Proxy in estimating grassland AGB with temporally optimized conditions.

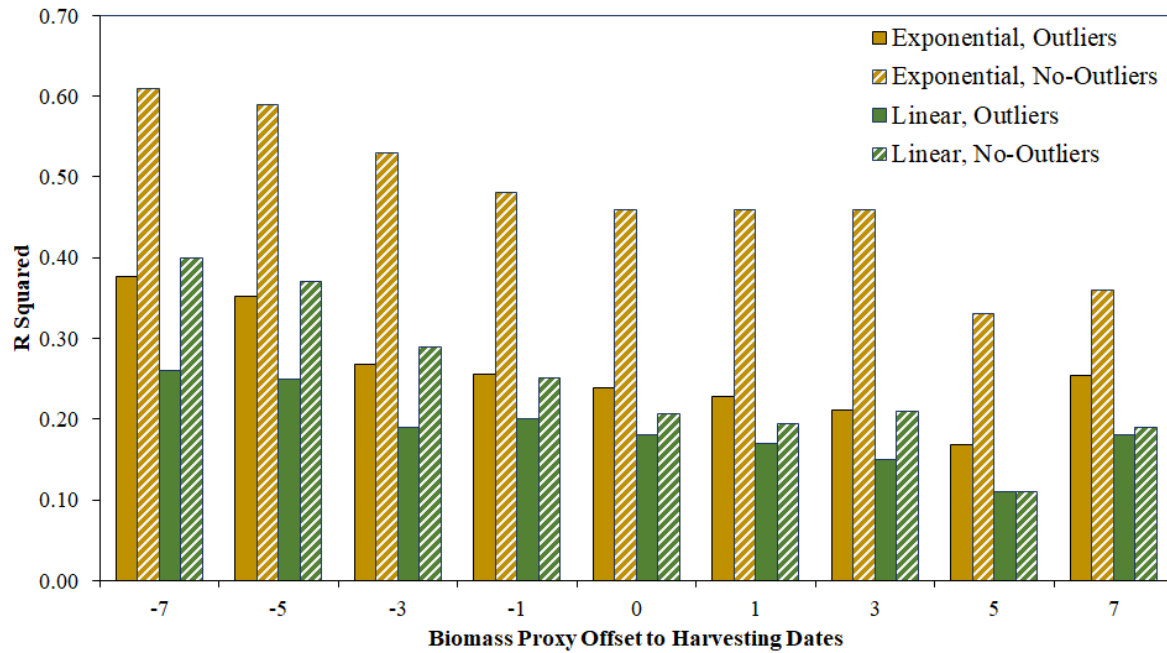


Figure 3.12. Coefficient of determination (R^2) for linear and exponential regression models between Biomass Proxy and in-situ data at different temporal offsets with/without outliers considered.

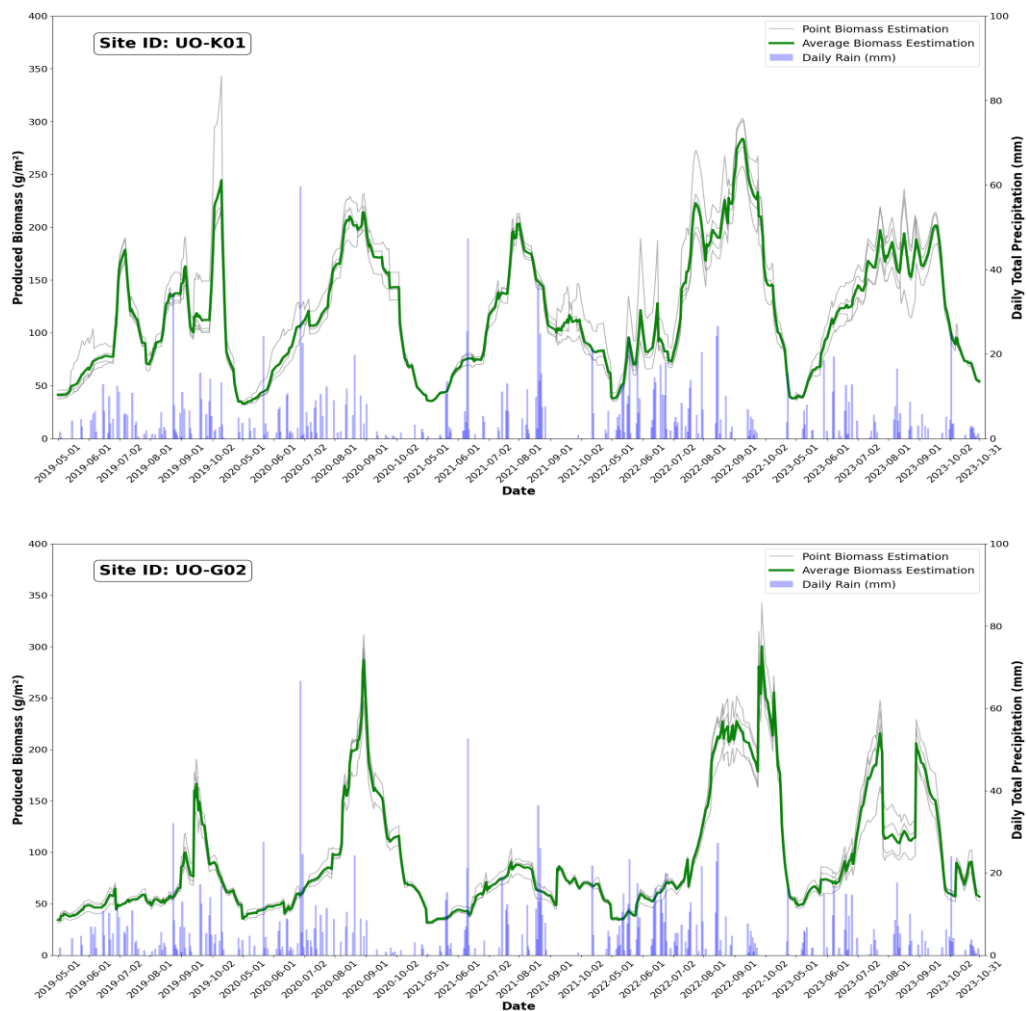
3.4.2. High-Temporal Biomass Prediction

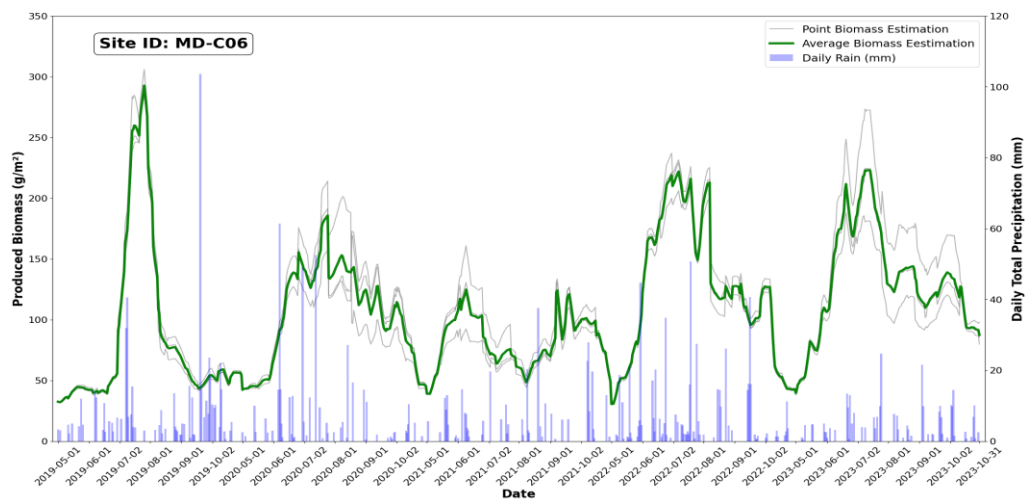
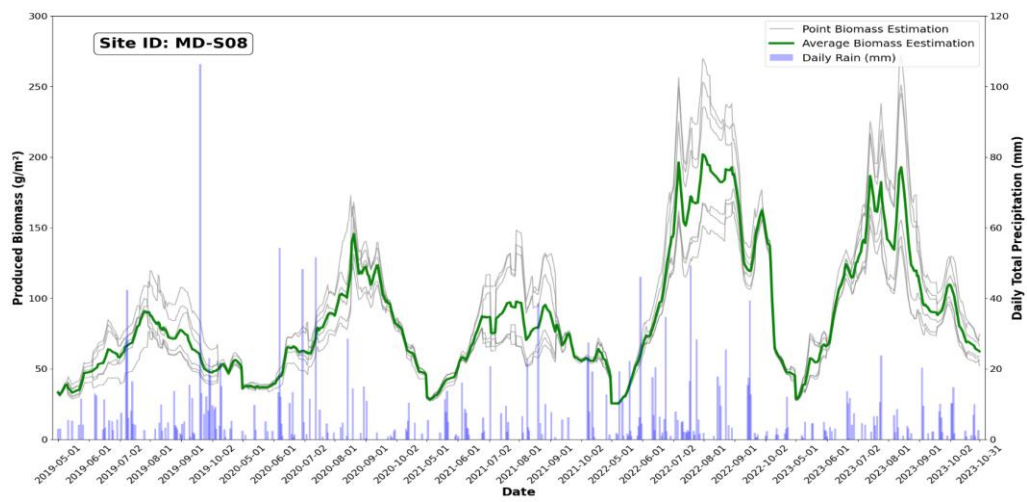
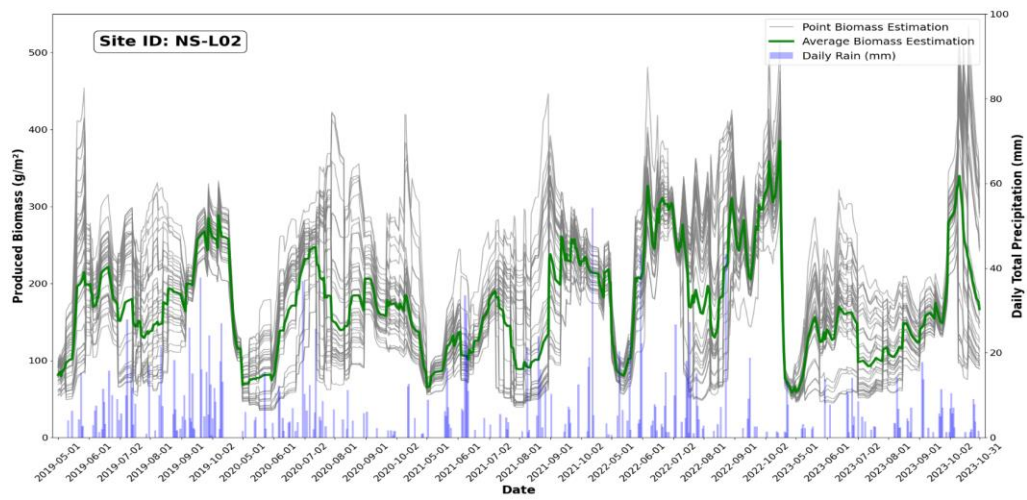
The limited in-situ biomass measurements impact the accuracy of developing a reliable prediction model using machine learning, which is highly needed to improve modelling generalization and stability. To fully assess grassland AGB within the study area throughout various years, the use of training data acquired during periods of growth in several years is necessary. Using the Biomass Proxy time-series data for the years ranging from 2019 to 2023, it was possible to generate new biomass samples consistent with satellite observations at various dates. Among linear and exponential regression models developed using Biomass Proxy variables, the exponential model with a seven-day offset from (i.e. prior to) the harvest date performed best and was thus used to produce synthetic biomass [Equation (12)]:

$$Biomass = 20.224 \times e^{3.58 \times (Biomass Proxy)} \quad (12)$$

Figure 3.13 shows the daily variation of synthetic biomass, calculated using Equation (12), across six farms, from UO-K01 in northwestern PE of Manitoba to farm MD-C06 in southeastern PE of Manitoba, during the growing season over five years. This study focused on the last three years of the period, i.e. 2021-2023, to estimate biomass, due to data

limitations. The green curve represents the average produced biomass of multiple locations shown by grey curves within each farm, as well as the average daily precipitation of the same locations. The plots showed that 2021 recorded the lowest peak biomass in all sites, highlighting the impact of that year's drought conditions. In contrast, the biomass peak comparisons between 2022 and 2023 varied by location, indicating that geography played a role in the peak biomass level. As shown in Figure 3.13, northern sites such as UO-K01 and UO-G02 exhibited higher biomass peaks in 2022 compared to both 2021 and 2023. In southern sites, including MD-S08, MD-N03, and MD-C06, there was little to no difference in biomass peaks between 2022 and 2023. In the NS-L02 site, between the northern and southern groups, the 2022 biomass peak was slightly higher than that of 2023.





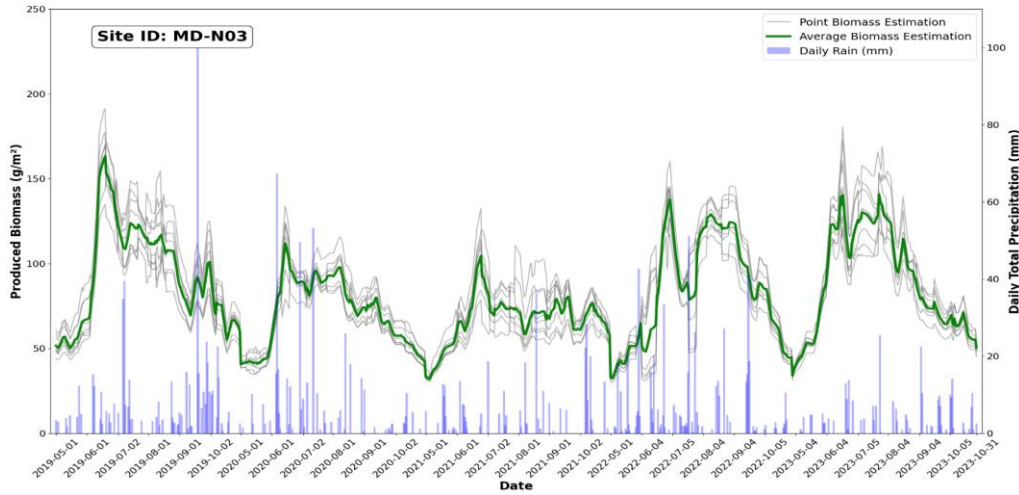


Figure 3.13. Temporal variation of synthesized biomass values during the growing seasons from 2019 to 2023 at the locations of all biomass samples across six fields (the site ID at the top left of each plot corresponds to the geographic position of the site as shown in Figure 3.3), produced using the Biomass Proxy predictor. Daily precipitation is also presented in the second Y-axis.

3.4.3. Grassland AGB Estimation Using Synthetic Samples

These produced grassland biomass samples were used to train the biomass models based on remote sensing features and environmental variables. Remote sensing features extracted from PlanetScope (3 m), Sentinel-2 (10 m) and Landsat 8-9 (30 m) were tested separately. The study period was limited to the years 2021–2023, as PlanetScope imagery was only purchased for these years, despite the sensor being operational from 2019 onward. Approximately 90 images were individually collected from every sensor within the 2021–2023 growing seasons over the target fields. The number of images can be considered as the number of field observations on different dates. Since continuous in-situ biomass measurements were generated using Biomass Proxy, it was possible to assign biomass samples to each observation with a temporal offset to the image acquisition date. Around 650 new biomass samples were produced for these observations.

3.4.3.1. Parametric Models by VIs and Environmental Variables

Figure 3.14 shows the heatmap of Pearson correlation (r) between grassland AGB with spectral bands of PlanetScope, VIs, elevation (DEM), precipitation (Prcp) and temperature (Temp) and correlations among the predictors themselves. The correlation coefficients showed that VIs predictors had the highest correlation with AGB (GLA, PanNDVI, and RVI had the three highest correlations with $r = 0.74$, 0.73 , and 0.71 , respectively), indicating that remote sensing features provided a good characterization of the grassland AGB. The

elevation feature had a positive correlation with grassland AGB, with $r = 0.47$, which indicated that grassland AGB at higher elevations increased. Among meteorological variables, the correlation of AGB with precipitation was higher than that with temperature ($r = 0.26$ and 0.23 , respectively).

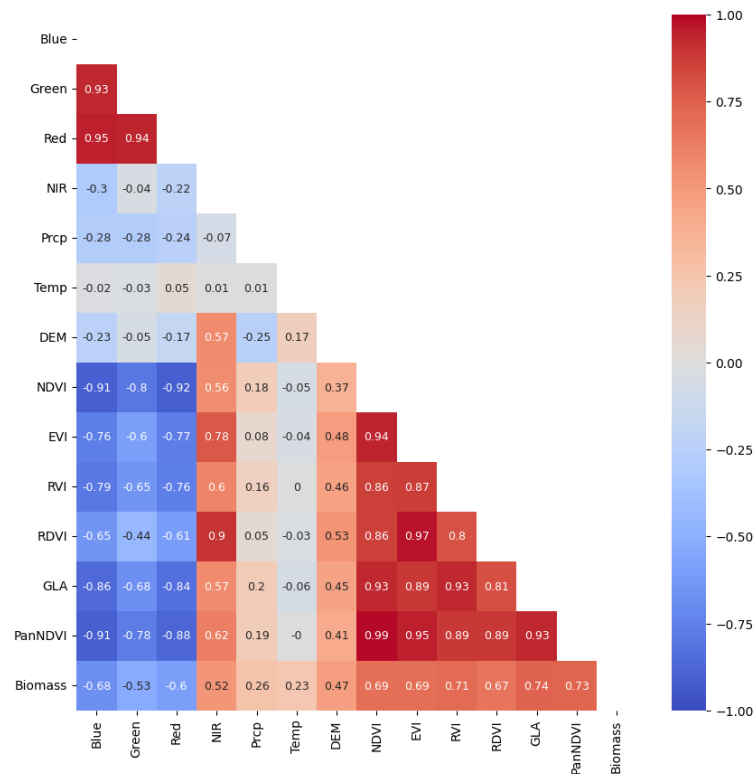


Figure 3.14. Pairwise Pearson correlation (r) between grassland AGB with spectral bands of PlanetScope, vegetation indices (VIs), elevation (DEM), precipitation (Prcp) and temperature (Temp).

According to correlation values between grassland AGB and the explanatory variables and correlation between predictors with themselves, a series of feature combinations were designed, and the results are shown in Table 3.6. Different feature groups were engineered due to their strong correlation with biomass and low inter-correlation between themselves (e.g., pairs of “PanNDVI, RDVI”, “NDVI, RDVI”, “NDVI, EVI” and “GLA, RDVI”). The impact of environmental variables was also tested by including and excluding them from models. The last model incorporated all features for AGB estimation. Regarding exponential regression models, only the models that included all features and those features that achieved the highest performance in linear models were tested.

The results demonstrated that the performance of the model increased with the addition of environmental variables. From a simple model using only PanNDVI and RDVI

($R^2=0.49$, $RMSE=51.82$), adding topographical (DEM) and climatic variables (precipitation and temperature) improved predictive efficiency ($R^2=0.53$, $RMSE=49.82$). Among the models testing different pairs of VIs alongside environmental variables, Model No. 9 (GLA, RDVI, DEM, precipitation, temperature) with $R^2=0.62$ and $RMSE=44.96$ indicated the best performance for biomass estimation. The best model (Model No. 10), which included all VIs and environmental variables, explained the biomass with $R^2=0.65$ and the lowest $RMSE$ of 43.24. Overall, these results emphasize the necessity of combining both spectral and environmental features to enhance a grassland AGB prediction. In exponential models, Model No. 11: (GLA, RDVI, DEM, precipitation, temperature) reached $R^2=0.60$ and $RMSE=45.94$, while for Model No. 12 (all features) $R^2=0.64$ and $RMSE=43.82$.

Table 3.6. The performance of linear regression with different combinations of predictors extracted from PlanetScope using synthesized biomass samples ($n=668$).

Models	Variables	R^2	RMSE
Linear			
No. 1	PanNDVI, RDVI	0.49	51.82
No. 2	PanNDVI, RDVI, DEM	0.53	49.82
No. 3	PanNDVI, RDVI, Prcp	0.51	51.04
No. 4	PanNDVI, RDVI, Prcp, Temp	0.57	47.68
No. 5	PanNDVI, RDVI, DEM, Prcp, Temp	0.61	45.75
No. 6	NDVI, RDVI	0.47	53.28
No. 7	NDVI, RDVI, DEM, Prcp, Temp	0.59	46.41
No. 8	NDVI, EVI, DEM, Prcp, Temp	0.59	46.49
No. 9	GLA, RDVI, DEM, Prcp, Temp	0.62	44.96
No. 10	NDVI, EVI, RVI, RDVI, GLA, PanNDVI, Prcp, Temp, DEM	0.65	43.24
Exponential			
No. 11	GLA, RDVI, DEM, Prcp, Temp	0.60	45.94
No. 12	NDVI, EVI, RVI, RDVI, GLA, PanNDVI, Prcp, Temp, DEM	0.64	43.82

According to the correlation matrix (Figure 3.15), VIs achieved the strongest positive correlations with grassland AGB among different predictor groups. Among the VIs, the highest correlations were observed with GLA ($r=0.70$), NDVI ($r=0.70$), RVI ($r=0.71$), and PanNDVI ($r=0.68$), all confirming the high relevancy value of these indices to biomass estimation. EVI ($r=0.53$) and RDVI ($r=0.45$) also made a moderate contribution. Besides the spectral indices, the environmental variables of DEM ($r=0.43$) and precipitation ($r=0.27$) were also moderately correlated with AGB, whereas the temperature ($r=0.05$) was weakly correlated. Such results demonstrate the importance of the use of VIs and some of the environmental characteristics for predicting the amount of grasslands, and that the thoughtful choice of informative

variables can contribute toward improving model accuracy, especially when using the Sentinel-2 data.

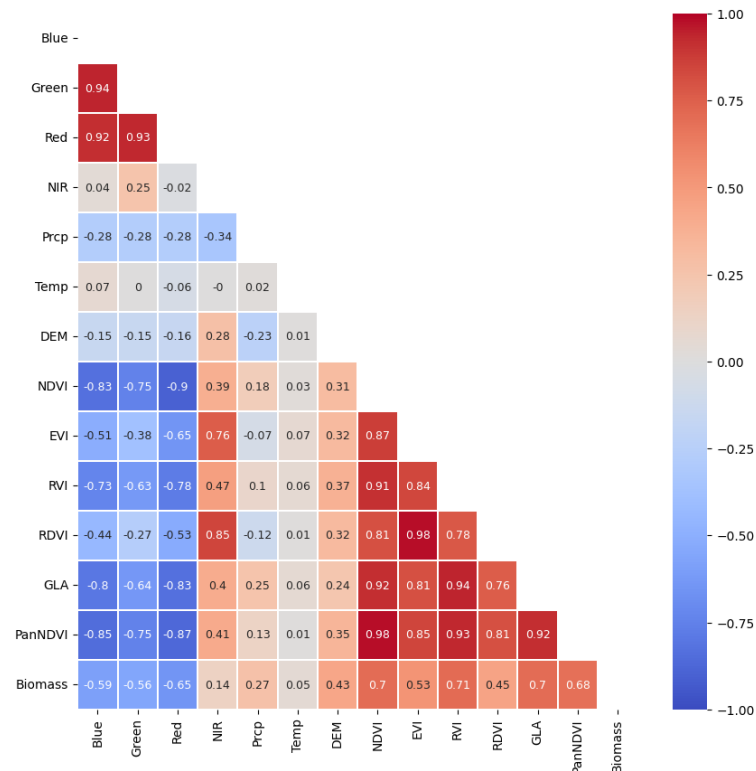


Figure 3.15. Pairwise Pearson correlation (r) between grassland AGB with spectral bands of Sentinel-2, vegetation indices (VIs), elevation (DEM), precipitation (Prcp) and temperature (Temp).

According to correlation analysis, some feature combinations, similar to the PlanetScope, were tested to estimate grassland AGB with different variables. Table 3.7 demonstrates a series of regression models based on VIs and environmental variables in the form of linear and exponential regression types. In the linear models, simpler combinations such as NDVI and EVI or RVI and EVI yielded only moderate results ($R^2=0.52$). However, including topographic (DEM) and climatic variables (precipitation and temperature) greatly enhanced the accuracy of the models; Model No. 6, which incorporated NDVI, EVI, Prcp, Temp and DEM, attained an R^2 of 0.60 and an RMSE of 46.44. The best linear model (Model No. 7) that used six VIs, as well as climatic and topographic variables, had the best accuracy with $R^2=0.64$ and RMSE=44.05. The exponential regression models demonstrated similar trends; Model No. 9 represented the best linear model at $R^2 = 0.65$, only with a slightly lower RMSE value.

Table 3.7. The performance of linear regression with different combinations of predictors extracted from Sentinel-2 using synthesized biomass samples (n=628).

Models	Variables	R ²	RMSE
Linear			
No. 1	NDVI, EVI	0.52	50.64
No. 2	RVI, EVI	0.52	50.62
No. 4	RVI, EVI, DEM	0.54	49.80
No. 5	RVI, EVI, Prcp, Temp	0.55	48.81
No. 6	RVI, EVI, Prcp, Temp, DEM	0.60	46.44
No. 7	NDVI, EVI, RVI, RDVI, GLA, PanNDVI, Prcp, Temp, DEM	0.64	44.05
Exponential			
No. 8	RVI, EVI, Prcp, Temp, DEM	0.55	49.24
No. 9	NDVI, EVI, RVI, RDVI, GLA, PanNDVI, Prcp, Temp, DEM	0.65	43.29

Finally, Landsat imagery with a spatial resolution of 30 m was used to estimate grassland AGB. Figure 3.16 illustrates the heatmap of Pearson correlation between the same VIs (derived from Landsat bands), climatic variables and DEM. The findings showed that the biomass values were strongly related to most features, with only a few exceptions. Notably, the intercorrelation among the VIs features themselves was comparatively low. Among the VIs, RVI ($r=0.75$) and RDVI ($r=0.71$) had the strongest correlations with biomass, whereas others, like EVI ($r=0.25$) and GLA ($r=0.63$), showed relatively weaker associations. Precipitation ($r=0.16$) and temperature ($r=0.09$) demonstrated weak correlation with AGB.

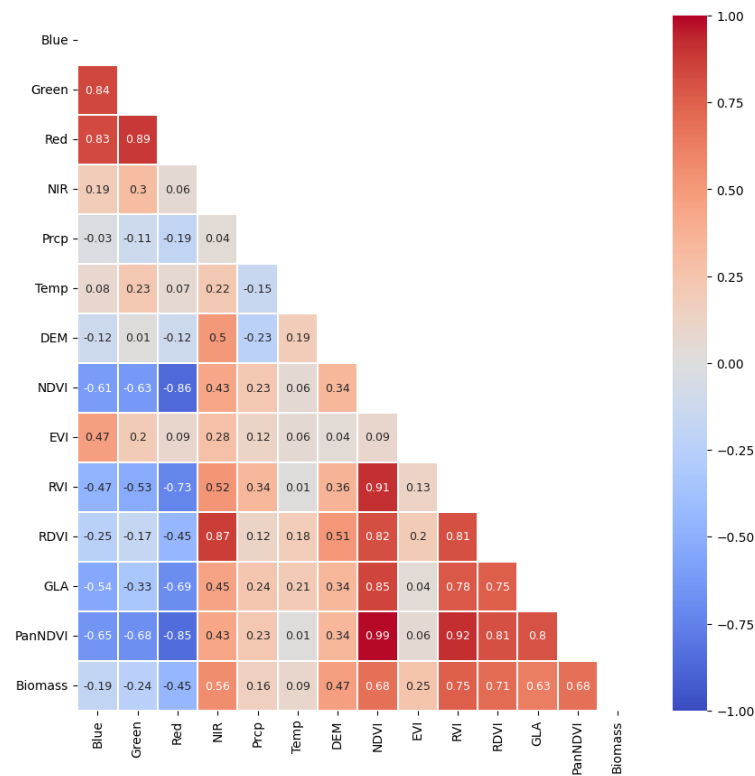


Figure 3.16. Pairwise Pearson correlation (r) between grassland AGB with spectral bands of Landsat 8-9, vegetation indices (VIs), elevation (DEM), precipitation (Prcp) and temperature (Temp).

Similar to previous parts, based on correlation analysis, some feature combinations were planned, and the results were provided in Table 3.8. The model's performance gradually increased with the further addition of environmental variables and VIs. The baseline model of RDVI and RVI (Model No. 1) in the linear models had an R^2 of 0.56. Incorporation of topographic information (DEM) in Model No. 2 marginally increased R^2 to 0.59, and again, the incorporation of climatic variables (precipitation and temperature) gave a moderate increase. The highest scoring linear model (Model No. 6) that used all six VIs (NDVI, EVI, RVI, RDVI, GLA and PanNDVI), precipitation, temperature, and DEM had an R^2 of 0.65 and the lowest RMSE of 44.88. In a similar way, exponential models were improving as the number of variables increased. The model with the complete set of predictors, Model No. 8, achieved the highest overall performance with $R^2=0.63$ and $RMSE=42.64$, which was slightly lower than its linear version. The residuals of linear and exponential models with all features across the three dataset types are demonstrated in Figure 3.A1 in the Appendix of this chapter.

Table 3.8. The performance of linear regression with different combinations of predictors extracted from Landsat 8-9 using synthesized biomass samples (n=668).

Models	Variables	R^2	RMSE
--------	-----------	-------	------

Linear			
No. 1	RDVI, RVI	0.56	46.46
No. 2	RDVI, RVI, DEM	0.59	44.84
No. 4	RDVI, RVI, Prcp, Temp	0.57	46.24
No. 5	RDVI, RVI, Prcp, Temp, DEM	0.60	44.83
No. 6	NDVI, EVI, RVI, RDVI, GLA, PanNDVI, Prcp, Temp, DEM	0.65	40.88
Exponential			
No. 7	RDVI, RVI, Prcp, Temp, DEM	0.50	49.83
No. 8	NDVI, EVI, RVI, RDVI, GLA, PanNDVI, Prcp, Temp, DEM	0.63	42.64

The coefficients of linear parametric models for three types of data were provided in Table 3.9. Coefficient values were calculated for either all predictors or a subset of predictors, according to their Pearson correlation, used as input variables in linear models. The finding showed that there was multicollinearity with VIs. For instance, the negative NDVI coefficient in Model No. 10 of PlanetScope or Model No. 6 of Landsat 8-9 was unexpected, as NDVI generally has a positive relationship with biomass. The findings also showed that feature selection based on Pearson correlation was effective in reducing redundancy among predictors. However, in Sentinel-2 models, multicollinearity was present, indicated by inconsistent coefficient signs, e.g., the negative sign of the EVI coefficient in Model No. 6, although they have a known positive correlation with biomass.

Table 3.9. Coefficient estimates of linear regression models for PlanetScope, Sentinel-2 and Landsat 8-9 datasets, based on either all predictors or a selected subset.

Coefficients	Variables	PlanetScope Models		Sentinel-2 Models		Landsat 8-9 Models	
		No. 9	No. 10	No. 6	No. 7	No. 5	No.6
β_1	NDVI		-779.7847		832.9711		-635.7270
β_2	EVI		433.6310	-73.1311	-459.7050		792.9314
β_3	RVI		-3.6873	17.1295	19.9653	12.41253	3.0094
β_4	RDVI	166.4121	-474.2372		276.3474	186.4609	-822.5910
β_5	GLA	304.2763	374.3171		32.6466		248.1179
β_6	PanNDVI		568.7491		-591.3950		366.4547
β_7	Prcp	0.47379	0.44048	0.604181	0.23249	0.03826	0.00022
β_8	Temp	3.9129	3.2679	-0.15852	0.00773	0.45158	0.03877
β_9	DEM	0.103246	0.09115	0.179682	0.19180	0.11435	0.14312

3.4.3.2. Machine Learning Models by VIs and Environmental Variables

Grassland AGB estimation using machine learning approaches demonstrated higher predictive performance than parametric approaches. Three machine learning algorithms of RF, SVR and ANN were used on datasets from three satellite sensors, such as PlanetScope, Sentinel-2, and Landsat 8-9 (Table 3.10). Across all model combinations, RF and SVR

consistently performed better than ANN in terms of R^2 and RMSE, though ANN demonstrated more stable generalization when cross-validation was used. Among the three sensors, Sentinel-2 and PlanetScope had better predictive performance than Landsat 8-9, as indicated by metric values, like R^2 and RMSE.

In the case of PlanetScope data ($n = 663$), the RF model performed the best with $R^2 = 0.88$ and $RMSE = 24.88$, followed by SVR ($R^2 = 0.83$) and ANN ($R^2 = 0.78$). ANN, however, exhibited the best generalization since its R^2 increased to 0.82 in 4-fold cross-validation. The elimination of climatic and topographic factors (Prpcp, Temp, and DEM) significantly reduced the performance of all three algorithms. The sharpest decrease in performance was observed in the SVR model with PlanetScope, with R^2 decreasing to 0.48 ($\Delta = -0.35$) and RMSE increasing to 52.66. These findings show the high sensitivity of environmental parameters on the accuracy of biomass prediction using PlanetScope images.

The best overall performance of models was observed in Sentinel-2 ($n = 628$). SVR obtained the best R^2 value (0.90) among all the sensors with the least RMSE of 21.73, whereas RF and ANN had R^2 values of 0.88 and 0.80, respectively. However, these models also experienced the effect of the omission of environmental predictors. For instance, SVR experienced a large drop from 0.90 to 0.60 and a significant increase in RMSE (21.73 to 43.64). Although Sentinel-2 showed the highest test set performance, its SVR model's K-fold cross-validation R^2 (0.85) was slightly lower than that of PlanetScope (0.80), indicating that the test accuracy could have been influenced by advantageous data partitioning and that model generalization was slightly less consistent. Landsat 8-9 ($n = 668$) showed slightly lower R^2 values than PlanetScope and Sentinel-2 but remained highly predictive. RF achieved an R^2 of 0.87, and ANN came second with 0.84, and SVR attained 0.82. Similar to the other sensors, removing precipitation, temperature, and elevation negatively affected performance. For example, RF's R^2 dropped from 0.87 to 0.72, and SVR declined from 0.82 to 0.59. For Landsat 8-9, all models attained lower k-fold R^2 , indicating that their test R^2 results were less reliable. The residuals of these nine models are demonstrated in Figure 3.A2 in the Appendix of this chapter.

Table 3.10. The performance of random forest (RF), support vector machine (SVR) and artificial neural network (ANN) models using various optical remote sensing sensors such as PlanetScope, Sentinel-2 and Landsat 8-9 in estimating grassland AGB.

Sensor	Models	Predictors	R ²	RMSE	R ² K-Fold (k=4)
PlanetScope (n = 663)	RF	All features	0.88	24.88	0.84
		All features without Prcp, Temp and DEM	0.60	45.97	0.64
	SVR	All features	0.83	30.12	0.80
		All features without Prcp, Temp and DEM	0.48	52.66	0.52
	ANN	All features	0.78	34.55	0.82
		All features without Prcp, Temp and DEM	0.55	48.73	0.61
Sentinel-2 (n = 628)	RF	All features	0.88	23.96	0.86
		All features without Prcp, Temp and DEM	0.71	37.11	0.68
	SVR	All features	0.90	21.73	0.85
		All features without Prcp, Temp and DEM	0.60	43.64	0.48
	ANN	All features	0.80	30.74	0.82
		All features without Prcp, Temp and DEM	0.69	38.18	0.66
Landsat 8-9 (n = 668)	RF	All features	0.87	25.35	0.81
		All features without Prcp, Temp and DEM	0.72	36.84	0.72
	SVR	All features	0.82	29.37	0.78
		All features without Prcp, Temp and DEM	0.59	44.86	0.54
	ANN	All features	0.84	27.12	0.80
		All features without Prcp, Temp and DEM	0.69	38.23	0.64

The other solution to address multicollinearity in a regression model with correlated predictors is to use PCA. In this study, the first *m* principal components were determined by the criterion of explaining 95% of total variance, as opposed to a fixed number of components. As shown in Figure 3.10 and Table 3.11, the PCA-based dimensionality reduction aligned closely with the correlation-based feature selection used in the parametric models. The derived linear regression models with PCA showed moderate predictive accuracy ($R^2 \approx 0.60$), which was comparable to the performance of linear models using selectively chosen VIs ($R^2 = 0.60$ - 0.62). These results indicate that, even though PCA can be used to enhance data structure and alleviate the problem of multicollinearity, the linear models might not be adequate to capture the complex and possibly non-linear nature of the relationships between remote sensing variables and grassland biomass. Machine learning methods were better at handling multicollinearity than parametric models, as they maintained high predictive performance even with correlated features. For instance, the R^2 of RF models was around 0.87, which was close to the performance of models using all original predictors. Different

sensors attained slightly different component counts based on their redundancy level among predictors. For example, Landsat only needed six to explain 95% of the variance. According to Table 3.11, the contribution of different components to explaining the target variable indicated that most of variability was captured by the first two components, while the last one accounted for the least.

Table 3.11. Principal components analysis (PCA) for the random forest (RF) and the linear regression models created using PlanetScope, Sentinel-2 and Landsat 8-9 data.

Data Type / Model	R ²	Number of PCA components	Explained variance ratio per component
PlanetScope / RF	0.87	4	[0.641 ; 0.141 ; 0.112 ; 0.054]
PlanetScope / Linear	0.61		
Sentinel-2 / RF	0.88	5	[0.612 ; 0.143 ; 0.111 ; 0.080 0.031]
Sentinel-2 / Linear	0.62		
Landsat 8-9 / RF	0.86	6	[0.518 ; 0.157 ; 0.115 ; 0.094 ; 0.059 ; 0.024]
Landsat 8-9 / Linear	0.60		

The importance of different features used in machine learning methods can be calculated using the MDI metric. Nine driving variables, including the climatic factors and optical variables [Figure 3.17 (a)] showed the greatest importance in the AGB estimation and jointly contributed 98%, whereas the topographical factor contributed 2%. Notably, among the nine driving variables, PanNDVI (42%) and GLA (16%) exhibited prominent importance and jointly contributed 55% of the spatial AGB variation. For the remaining driving variables, Temp and Prcp had a relatively high importance rating of 15% and 8%, respectively. Based on the importance ratings of predictors within each specific category [Figure 3.17 (b)], VIs, climatic variables and DEM contributed 75%, 23%, and 2%, respectively, to the estimation of AGB.

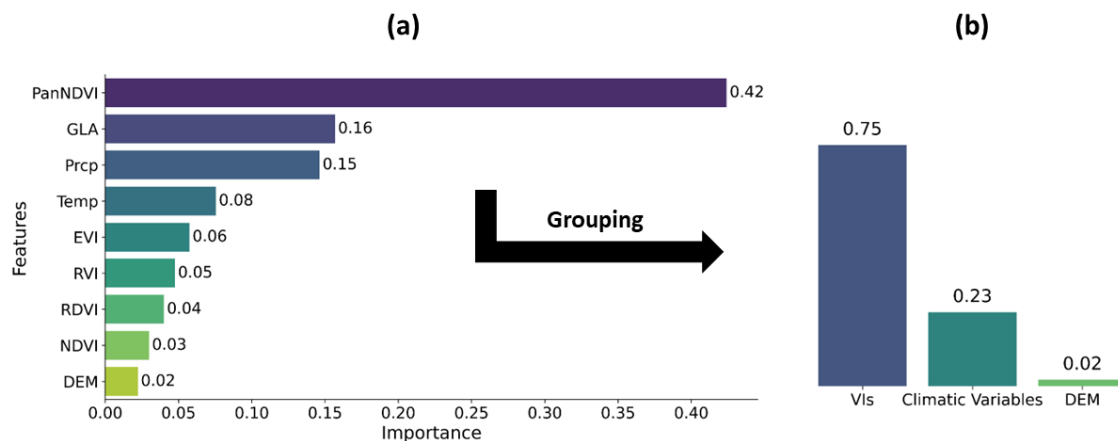


Figure 3.17. Predictor importance of modelling of grassland AGB across six fields in the Prairie Ecozone (PE). (a) Importance of nine driving variables individually and (b) within categories.

The relationship between observed and estimated AGB of different models is shown in Figure 3.18 and Figure 3.19 using scatter plots, with the 1:1 reference line (red dashed), indicating perfect agreement. Generally machine learning models demonstrated tighter clustering around the 1:1 line than parametric models. Parametric models tend to show more systematic bias, meaning stronger underestimation or overestimation than machine learning algorithms. Among all scatter plots, PlanetScope data and SVR and RF models exhibited a tighter clustering around the 1:1 line, indicating precision and consistency of AGB estimates as compared to Sentinel-2 and Landsat 8-9.

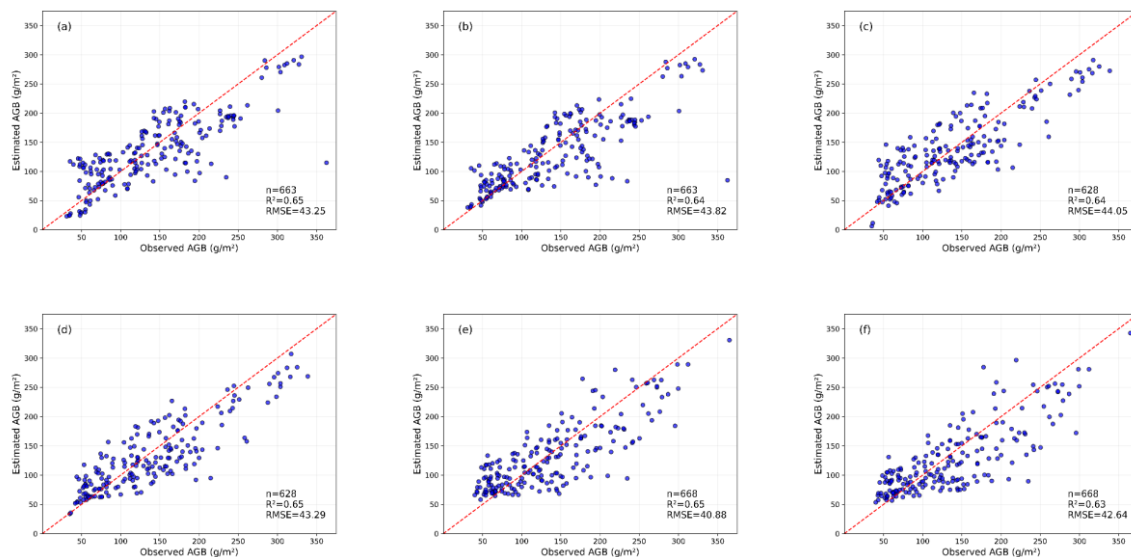
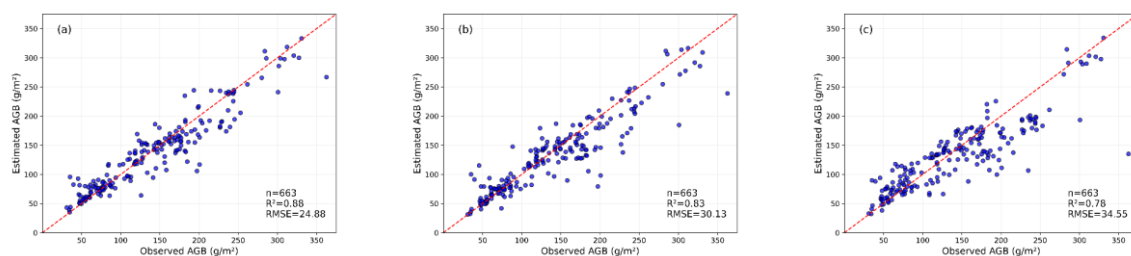


Figure 3.18. Relationship between observed and estimated AGB using parametric (linear and exponential) models based on PlanetScope, Sentinel-2 and Landsat 8-9 optical data: (a) linear-PlanetScope, (b) exponential-PlanetScope, (c) linear-Sentinel-2, (d) exponential-Sentinel-2, (e) linear-Landsat 8-9, and (f) exponential-Landsat 8-9.



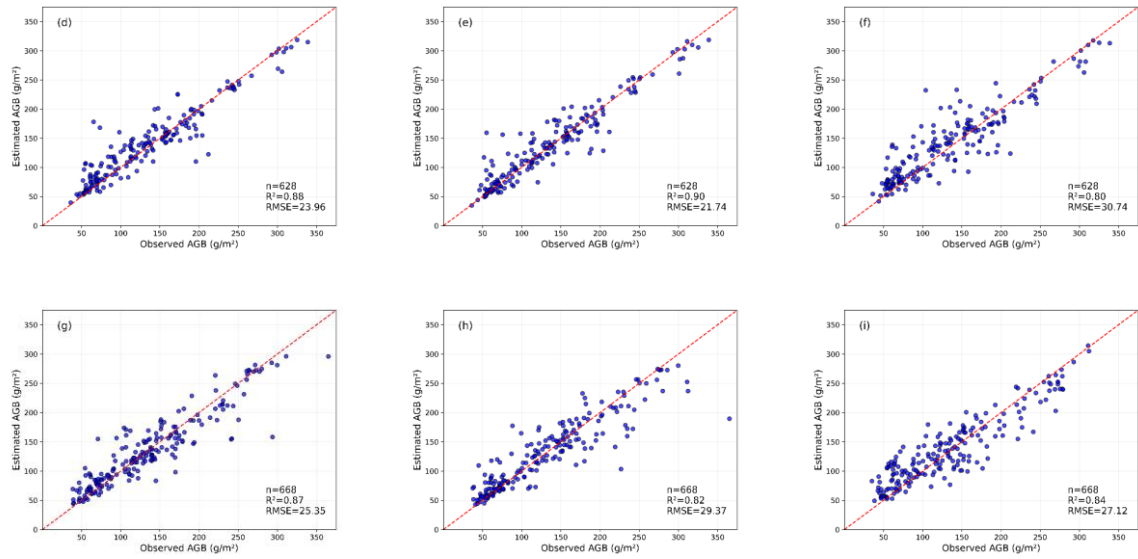
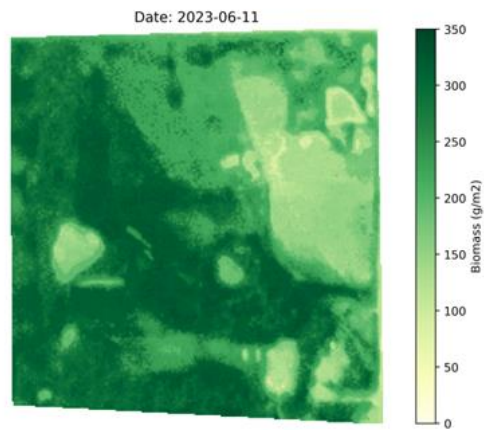


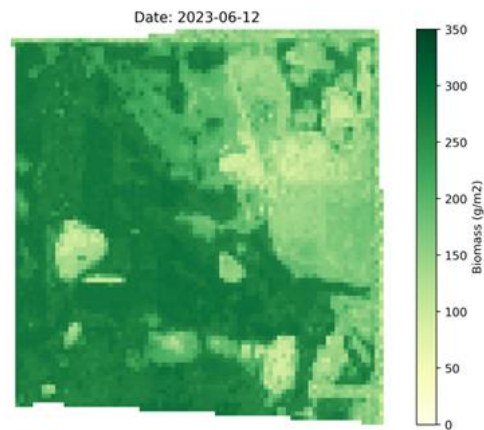
Figure 3.19. Relationship between observed and estimated AGB using machine learning [random forest (RF), support vector regression (SVR) and artificial neural network (ANN)] models based on PlanetScope, Sentinel-2 and Landsat 8-9 optical data: (a) RF-PlanetScope, (b) SVR-PlanetScope, (c) ANN-PlanetScope, (d) RF-Sentinel-2, (e) SVR-Sentinel-2, (f) ANN-Sentinel-2, (d) RF-Landsat 8-9, (e) SVR-Landsat 8-9, and (f) ANN-Landsat 8-9.

The study used RF as a robust and high-performing model across all satellite datasets to generate maps of grassland AGB at the field level. In Figure 3.20, AGB maps based on three sensors, PlanetScope, Sentinel-2, and Landsat 8-9, are shown around the same date to make it possible to visually compare the spatial resolution and biomass distribution. The finer resolution of PlanetScope allows detailed spatial patterns of biomass distribution, while coarser resolutions lead to increasingly generalized representations.

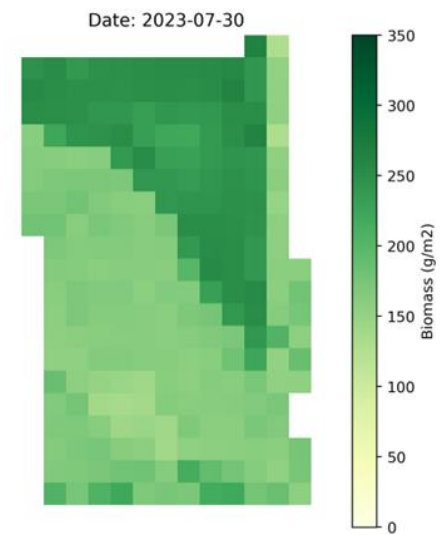
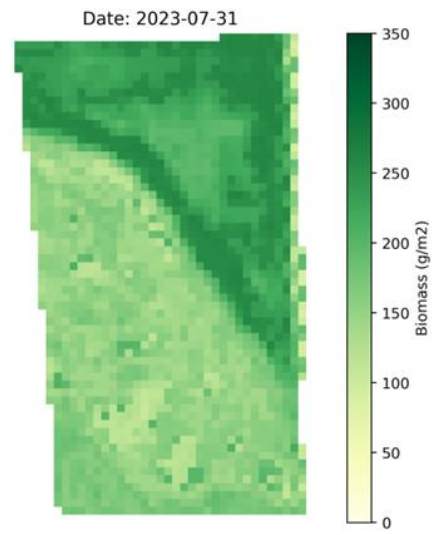
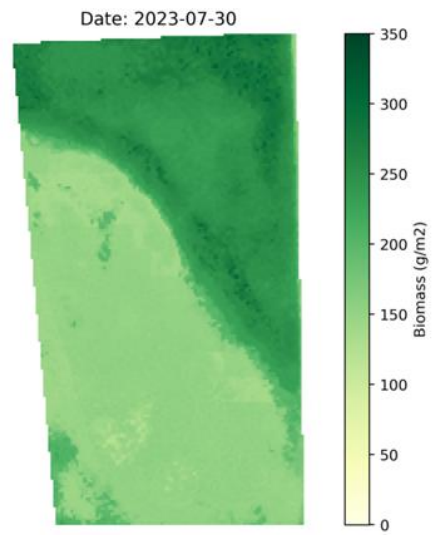
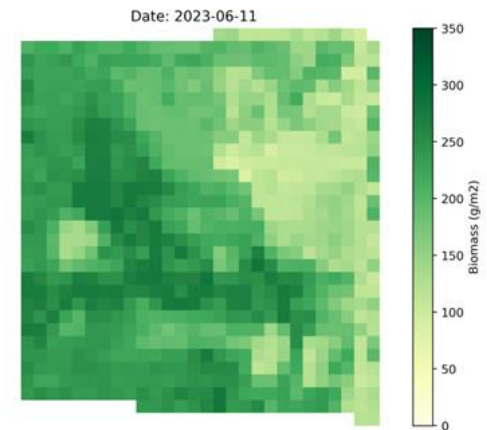
(a)

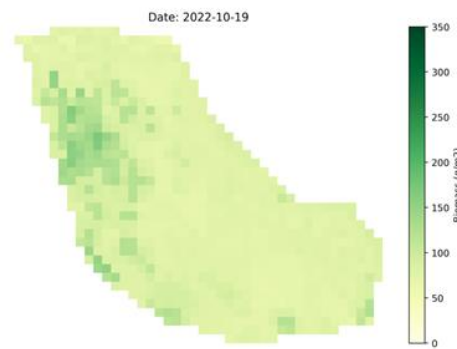
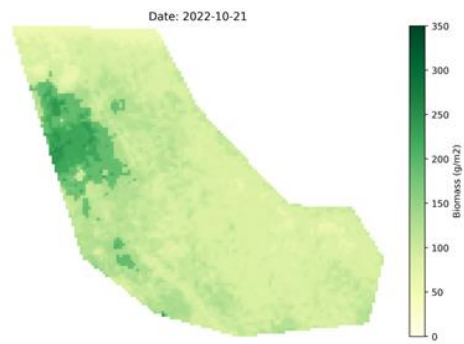
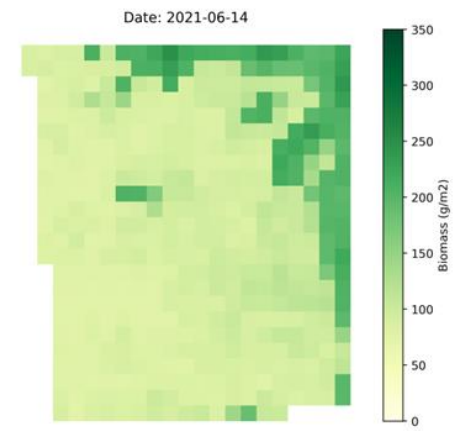
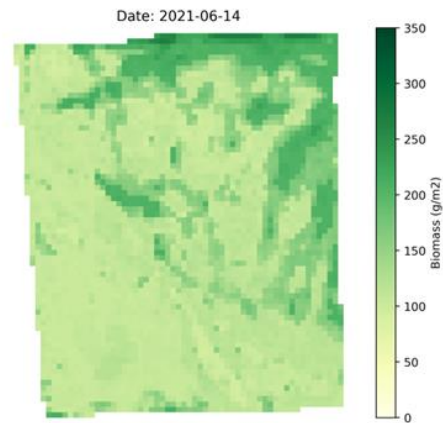
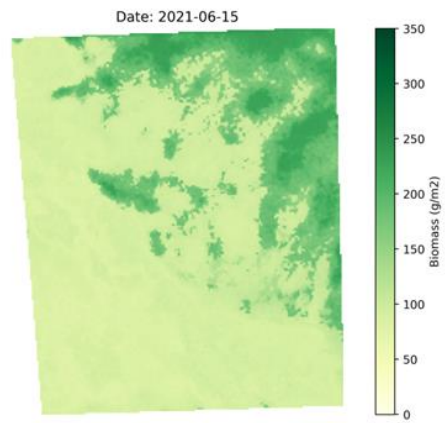


(b)



(c)





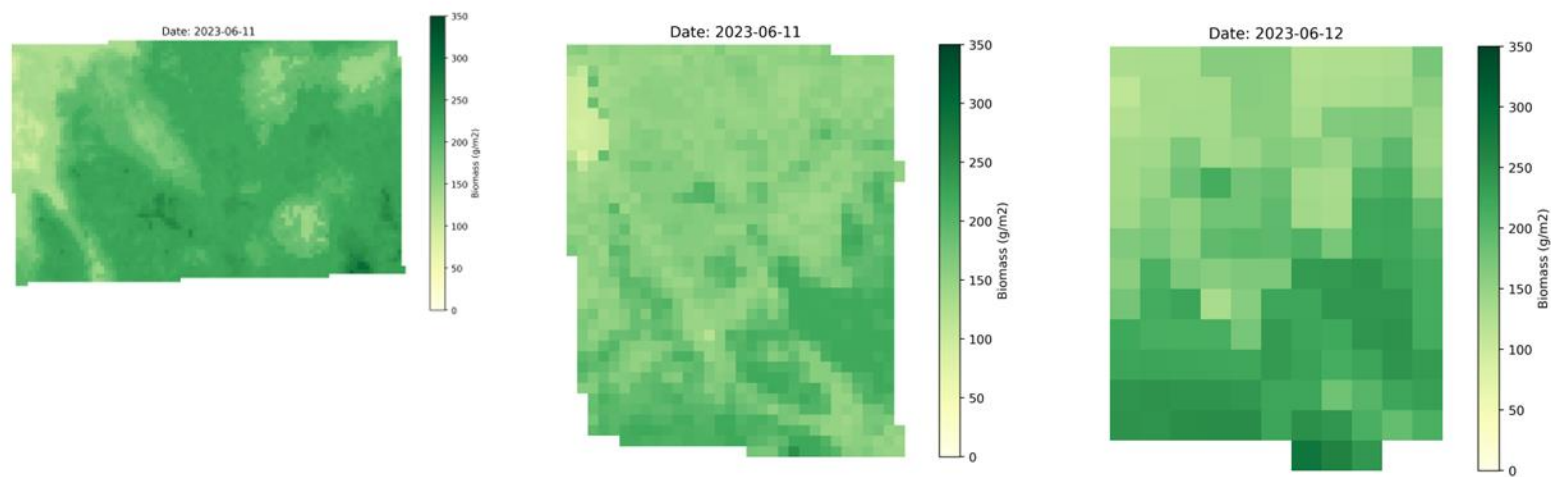


Figure 3.20. Maps of grassland AGB estimated using the random forest (RF) model with input data from (a) PlanetScope, (b) Sentinel-2, and (c) Landsat 8-9, acquired at different locations and dates.

3.5. Discussion

This chapter assessed the capacity of various remote sensing data and regression models for estimating grassland AGB in the Prairie region of Manitoba. The study examined the contribution and effectiveness of six well-known VIs, climatic variables (e.g., precipitation and temperature), and elevation in estimating grassland AGB by incorporating them as input features in both parametric and non-parametric predictive models. Several important observations emerged from the analysis, which are discussed in the following sections.

Before removing outliers, the R-squared values for both linear and exponential models were low, which indicated that the vegetation index predictors were weak in predicting biomass. However, the predictors were statistically significant ($p < 0.001$). This could be because noisy observations were used in univariate regression models. The removal of outliers significantly improved the predictive accuracy of predictors, particularly NDVI, RDVI and PanNDVI exponential models. Univariate analysis revealed that exponential models were better than linear models in estimating biomass using most vegetation index predictors, suggesting a non-linear relationship between spectral indices and AGB. Nevertheless, some predictors, like RVI and GLA, had linear behaviour regarding biomass. The RVI-based linear model achieved the highest R^2 among other VIs, either with or without outliers. These findings, which highlight the effective role of the NDVI predictor for modelling biomass, is congruent with previous reviews, which performed well in biomass estimation. Gao et al. (2013) developed unitary linear, exponential and power regression models using MODIS-NDVI, indicating the exponential model with R^2 of 0.66 achieved the best performance. F. Li et al. (2013) reported that the univariate linear regression model with NDVI predictor (with $R^2=0.50$) outperformed EVI ($R^2=0.32$) and soil adjusted vegetation index (MSAVI) ($R^2=0.31$), while linear model with RVI and land-surface water index (LSWI) predictors achieved slightly higher performance ($R^2 = 0.50$ and 0.54 , respectively).

It was also discovered that the Planet Biomass Proxy was validated as a reliable product for biomass estimation. All VI-based models were outperformed by exponential regression using Biomass Proxy data shifted by seven days prior to biomass sampling, which yielded the highest R^2 value (0.61). This could be because the property of the Biomass Proxy, showing a fused index of SAR and optical data, was superior at tracking vegetation dynamics. Due to the freshness of the Biomass Proxy product, scholars have not yet assessed effectively the effects

of the product on grassland AGB. Guillevic et al. (2024) evaluated the relationship between Biomass Proxy and fresh biomass of corn and soybean, showing a strong correlation between crop fresh biomass and Biomass Proxy ($R^2 = 0.90$ and 0.68 , respectively).

Extracting the spectral indices corresponding to in-situ biomass measurements had some obstacles due to atmospheric conditions (e.g., cloud and aerosol) and temporal resolution that reduced the observations capability. These limitations avoid having a consistent temporal interval between all remote sensing data (i.e., VIs) and biomass samples. For example, in the study, the temporal distance ranged between 1 and 11 days. On the other hand, Planet Biomass Proxy, which offers a daily, cloud-free, analysis-ready product, allows users to select any temporal interval between Biomass Proxy and in-situ measurements, such that the interval remains uniform for all observations. The findings showed that if the Biomass Proxy product had an interval of seven days with in-situ measurements, it delivered the highest correlation with in-situ data. This may be due to inconsistency between the pixel size on the ground and the in-situ biomass sample size, which prevents the pixel value of Biomass Proxy from properly representing the in-situ sample. Biomass Proxy also suffers from saturation due to the use of optical index in its structure (Guillevic et al., 2024); thus, obtaining data a few days before peak biomass might decrease the saturation effects and build a stronger relationship.

To fully understand the dynamics of grassland biomass, it is essential to collect in-situ data throughout the entire growing season, which represents the grassland AGB during this period. However, collecting grassland samples over several years and multiple locations is expensive and time-consuming. For training and validation of machine learning models and having robust and reliable results, the availability of comprehensive, seasonally representative ground-truth data is critical. In addition, multi-year data let the research determine inter-annual variability in AGB owing to environmental drivers. Transfer learning (Pan & Yang, 2010), data augmentation (Shorten & Khoshgoftaar, 2019), simpler models (Xu et al., 2023), and regularization techniques (Tian & Zhang, 2022) are solutions to the issues faced by limited data. Generating synthetic biomass data using Biomass Proxy from its strong relationship with in-situ biomass samples allows us to populate biomass samples for a longer period (2019-2023) at different locations. Climatic variables such as temperature and precipitation from the studied days/years improve the multi-temporal AGB prediction by

capturing how precipitation influences biomass dynamics, which is vital for detecting inter-annual variability in AGB in response to climate (Ali et al., 2017). Although synthetic data were not entirely like real data because they were the result of a regression that explained 61% of the real biomass variation, they were qualitatively similar to real data, covering the dynamics of grassland biomass, such as seasonal growth patterns and interannual variability.

The multivariate parametric models' results using synthetic biomass samples indicated that when all features, such as VIs and environmental factors, were included, the overall performance of the models was improved. In multivariate linear regression models, Landsat 8-9 and PlanetScope showed a slightly higher performance ($R^2=0.65$) than Sentinel-2 ($R^2=0.64$). In multivariate exponential regression, however, the trend was the opposite, with Sentinel-2 returning the best results ($R^2=0.65$), with Landsat 8-9 and PlanetScope close behind ($R^2=0.64$). When selected predictors were used, linear models outperformed exponential models for all satellite data, showing a 0.02 to 0.10 unit difference in R^2 . This is likely because the variable selection was based on Pearson matrix information, which reflects linear relationships and is better aligned with linear regression modelling. In parametric regression models, multicollinearity occurs when predictors are highly correlated with each other, which can affect the reliability and interpretation of the regression results. Multicollinearity inflates the standard error of coefficient estimates, thereby reducing the reliability of the model. The other consequences are the sensitivity of the coefficient estimates to small changes, and difficulties in interpreting the fitted coefficient of each independent variable (Montgomery et al., 2021; Neter et al., 1983). To avoid multicollinearity, the predictors that showed a high correlation with biomass and did not exhibit high correlations among themselves were selected based on their correlation matrix. If variables had a Pearson correlation above 0.9, they were assumed to share overlapping information and were thus considered collinear. For instance, with PlanetScope data, several variable selections were designed, such as pairs of "PanNDVI and RDVI", "NDVI and RDVI", "GLA and RDVI" and "NDVI and EVI" alongside the environmental features, that combination of GLA and RDVI and environmental features achieved the highest performance ($R^2=0.62$). The same strategy was applied to Sentinel-2 and Landsat 8-9 data to build regression models with approximately independent variables. Predictors "RVI and EVI", and "RDVI and RVI", extracted from Sentinel-2 and Landsat 8-9 data, respectively, alongside the environmental features, were selected for further regression

analysis. The PCA dimensionality reduction technique was also used alongside the correlation-based feature selection approaches to handle correlated features in linear regression. The results showed that the average R^2 of PCA methods (≈ 0.61) supported the results of the correlation-based feature reduction method (≈ 0.61).

Although parametric models offered valuable insights, they struggled to handle the complex and non-linear relationships present in the data. Ge et al. (2018) demonstrated that SVR regression model outperformed multivariate parametric models such as linear and exponential regressions. Ali et al. (2017) compared parametric models to machine learning models using five VIs and two raw bands (Red, NIR) extracted from MODIS products, and the results indicated that ANN ($R^2=0.63, 0.59$) and adaptive neuro-fuzzy inference system (ANFIS) ($R^2=0.85, 0.76$) produced better estimations of biomass compared to multiple linear regression ($R^2=0.39, 0.29$) at two locations. The findings of this study also supported the previous research; based on Table 3.10 and Table 3.6-3.8, machine learning methods developed a higher performance in biomass estimation with lower error. RF method showed constant high performance across three sensors' datasets (R^2 ranging from 0.87 to 0.88 and RMSE ranging from 23.96 to 25.35). However, SVR and ANN had more fluctuations in their results among different data types. SVR was also effective and maintained its reliability even when the source of input features changed. With Sentinel-2 data, SVR models achieved R^2 values as high as 0.90, highlighting their strong predictive capability. A comparative study was conducted by Qin et al. (2021) to compare VI models for estimating fresh AGB using Sentinel-2 and Landsat. Their findings showed that generally, VI models using Sentinel-2 data built stronger relationships with AGB and achieved higher R^2 compared to those using Landsat 8. Their results may also be extended to VI models using PlanetScope, as its spectral bands (i.e., Blue, Green, Red and NIR) in the new generation satellite Cubes are specially designed to match the spectral bands of Sentinel-2 (PlanetScope | Planet Documentation, 2025). ANN did moderately well, but their R^2 values were generally lower than the other algorithms, likely because ANN requires more data and careful tuning of hyperparameters (Bergstra & Bengio, 2012). When all features were included (i.e., VIs and environmental variables), the ANN model with Landsat 8-9 data achieved better test-set results compared to PlanetScope and Sentinel-2 data. This might be attributed to the lucky partitioning of training and test sets for Landsat data and less favourable splits for the others. For instance, the k-fold cross-validation revealed

that the averaged performance of the ANN model with Sentinel-2 data was slightly lower than with other sensors (0.80 vs. 0.82). Similarly, the strong performance of the SVR model with Sentinel-2 ($R^2=0.90$ on the test set) may have been overestimated, as its k-fold result was notably lower ($R^2=0.85$), indicating less stable generalization across data splits. Meanwhile, the study experienced no significant changes due to various spatial resolutions of satellite data in biomass estimation.

The AGB prediction improved when incorporating climatic variables such as precipitation and temperature, as well as elevation data, alongside VIs; this indicated that VIs alone were insufficient for capturing the full spectrum of biomass changes. This prediction performance was observed in both parametric and machine learning regression models. In parametric models, although correlation values between biomass samples and climatic variables varied due to differences in image acquisition times, the relationship direction remained positive, with low magnitude when temperature correlations ranged from 0.05 to 0.23, and precipitation correlations ranged from 0.16 to 0.27, and this can be distinct for different grassland types and locations. Wu et al. (2023) found that average grassland (AGB) in temperate grasslands of China increased with an increase in mean annual precipitation (MAP) and decreased with a rise in mean annual temperature (MAT). In alpine grasslands, there were strong positive correlations between AGB and MAP as well as MAT. In PlanetScope Model No. 9 and Landsat 8-9 Model No. 5, both of which had a relatively low multicollinearity and could be interpreted, the elevation coefficient was approximately 0.10, whereas the temperature coefficient was about ten times larger than that of precipitation. This implies that in these models, temperature variation can influence biomass change more than precipitation, while precipitation has a higher individual correlation with biomass. In machine learning models, after adding all the features, the R^2 values for RF, SVR and ANN experienced an increase of 0.28, 0.35 and 0.23 for PlanetScope data, which were greater than those of their corresponding methods with other sensors. Removing precipitation, temperature, and DEM caused the R^2 value to decrease from 0.87 to 0.29 in the Landsat 8-9-based model and from 0.90 to 0.60 in the Sentinel-2-based model. Because of these declines, it was apparent that these factors were valuable for the environment and could explain the distribution of biomass related to climate change and landforms. The impact on these features changed by satellite, as image dates and the way data were matched with the environment were not the

same everywhere. Thus, studying multiple types of environmental data with remote sensing helped strengthen remote sensing features for biomass interpretations and to obtain a deeper ecological understanding. The importance of the driving variables, both individual and category, is demonstrated in Figure 3.17, showing that environmental variables explain 25 percent of the grassland AGB. Among them, temperature and precipitation contributed 15% and 8%, respectively, while the importance rate of DEM was 2%. However, the highest contribution was dedicated to VIs predictors, with an importance value of 75%. The importance values of diverse VIs were not the same; some of which, such as PanNDVI (42%) and GLA (16%), showed higher importance, while RDVI (4%) and NDVI (3%) contributed less.

3.6. Conclusion

This research aimed to introduce a methodology for grassland biomass estimation using various remote sensing data, including PlanetScope, Sentinel-2 and Landsat 8-9 and regression models such as parametric and machine learning models. Firstly, various univariate linear and exponential regressions between a limited number of in-situ data and optical predictors like VIs and Biomass Proxy were made. Among VIs models, NDVI exponential regression achieved the best performance with R^2 of 0.53. However, Biomass Proxy, the combination of radar and optical indices, showed a higher R^2 (0.61). Outlier removal using Cook's distance stopped influential points from negatively affecting the regression models. Its effect on NDVI and Biomass Proxy were a rise of 0.16 and 0.22 in R^2 .

In the second stage, the synthetic biomass samples were used to build multivariate parametric and machine learning models to predict biomass. VIs and environmental features across 2021-2023 were used as input variables in parametric models. When all variables were involved, three sensor types showed approximately similar performance for linear and exponential regression with R^2 around 0.64. However, by removing some correlated VIs features, linear regression ($R^2=0.60-0.62$) outperformed exponential regression ($R^2=0.50-0.60$).

Machine learning models with the same features showed better results than parametric models. SVR and RF got the highest accuracy among all the sensors and were able to reduce RMSE, reaching a maximum R^2 of 0.90 and a minimum RMSE of 21.73. The RF method performed very well across different data types, and the PCA results indicated that

the RF models were effective in handling multicollinearity. It was also observed that environmental variables (temperature, precipitation and elevation) were effective in machine learning models' performance and removing them resulted in significant drops in R^2 and increases in RMSE, which confirmed these variables are key to estimating AGB. Meanwhile, it was found through feature importance that different VIs played varying roles. Both NDVI and PanNDVI (together 58% importance) were key components, whereas RDVI and RVI (together 7% importance) played a small role, meaning that choosing the right VIs by considering their spectral features and the attributes of the data is necessary to build solid models.

3.7. Appendix

Table 3.A1. The Pearson (r) and Spearman (ρ) correlations between biomass samples (n =102) and various vegetation indices (VIs) extracted from Sentinel-2 sensors.

Index	Formula*	Pearson correlation (r)	Spearman correlation (ρ)
ARVI	$(NIR - 2 \times R + B) / (NIR + 2 \times R - B)$	0.47	0.55
BGND	$(G - B) / (G + B)$	0.47	0.36
BNDVI	$(NIR - B) / (NIR + B)$	0.50	0.52
Cigreen	$CIG = RE4 / G - 1$	0.54	0.52
Cirededge	$RE4 / RE1 - 1$	0.48	0.49
Clg	$RE1 / G - 1$	0.07	0.00
Clr	$RE3 / RE1 - 1$	0.48	0.48
DVI	$NIR - R$	0.46	0.47
EVI	$2.5 \times ((NIR - R) / (NIR + 6 \times R - 7.5 \times B + 1))$	0.45	0.46
EVI2	$2.5 \times (NIR - R) / (1 + NIR + 2.4 \times R)$	0.48	0.49
ExB	$1.4 \times B - G$	-0.42	-0.35
ExG	$2 \times G - R - B$	0.42	0.49
ExGR	$ExG - ExR$	0.43	0.53
ExR	$1.4 \times R - G$	-0.42	-0.53
G_R	$G - R$	0.41	0.51
GBNDVI	$(NIR - (G + B)) / (NIR + (G + B))$	0.50	0.53
GLA	$(2 \times G - R - B) / (2 \times G + R + B)$	0.54	0.55
GNDVI	$(RE4 - G) / (RE4 + G)$	0.47	0.52
GRNDVI	$(NIR - (G + R)) / (NIR + (G + R))$	0.49	0.55
GRVI	G / R	0.54	0.57
IPVI	$NIR / (NIR + R)$	0.49	0.50
MGRDI	$(G^2 - R^2) / (G^2 + R^2)$	0.52	0.57
MSAVI	$(1/2) \times (2 \times (RE4 + 1) - \sqrt{(2 \times RE4 + 1)^2 - 8 \times (RE4 - R)})$	0.49	0.51
MTCI	$(RE2 - RE1) / (RE1 - R)$	0.26	0.25
NDVI	$(NIR - R) / (NIR + R)$	0.47	0.54
NDVire1	$(NIR - RE1) / (NIR + RE1)$	0.42	0.49
NDVire2	$(NIR - RE2) / (NIR + RE2)$	0.24	0.23
NDVire3	$(NIR - RE3) / (NIR + RE3)$	-0.23	-0.29
NGRDI	$(G - R) / (G + R)$	0.53	0.57
NIR_G	NIR / G	0.54	0.52
NIRv	$(NIR - R) / (NIR + R) \times NIR$	0.48	0.49
OSAVI	$(NIR - R) / (NIR + R + 0.16)$	0.48	0.53
OSAVI2	$1.16 \times (NIR - R) / (NIR + R + 0.16)$	0.48	0.53
PanNDVI	$(NIR - (R + G + B)) / (NIR + (R + G + B))$	0.51	0.55
PSRI	$(NIR - RED) / \sqrt{NIR + RED}$	-0.38	-0.49
RBNDVI	$(NIR - (R + B)) / (NIR + (R + B))$	0.50	0.55
RDVI	$(NIR - R) / \sqrt{NIR + R}$	0.48	0.50
RGBVI	$(G^2 - B \times R) / (G^2 + B \times R)$	0.54	0.55
RVI	NIR / R	0.58	0.54
RVI2	R / NIR	-0.44	-0.54
SAVI	$[(NIR - R) / (NIR + R + 0.5)] \times (1 + 0.5)$	0.48	0.50
TVI	$0.5 \times (120 \times (NIR - G) - 200 \times (R - G))$	0.46	0.47
VARI	$(G - R) / (G + R - B)$	0.52	0.56

* RE, NIR, R, B, G stands for Red Edge, near infrared, Red, Blue, Green

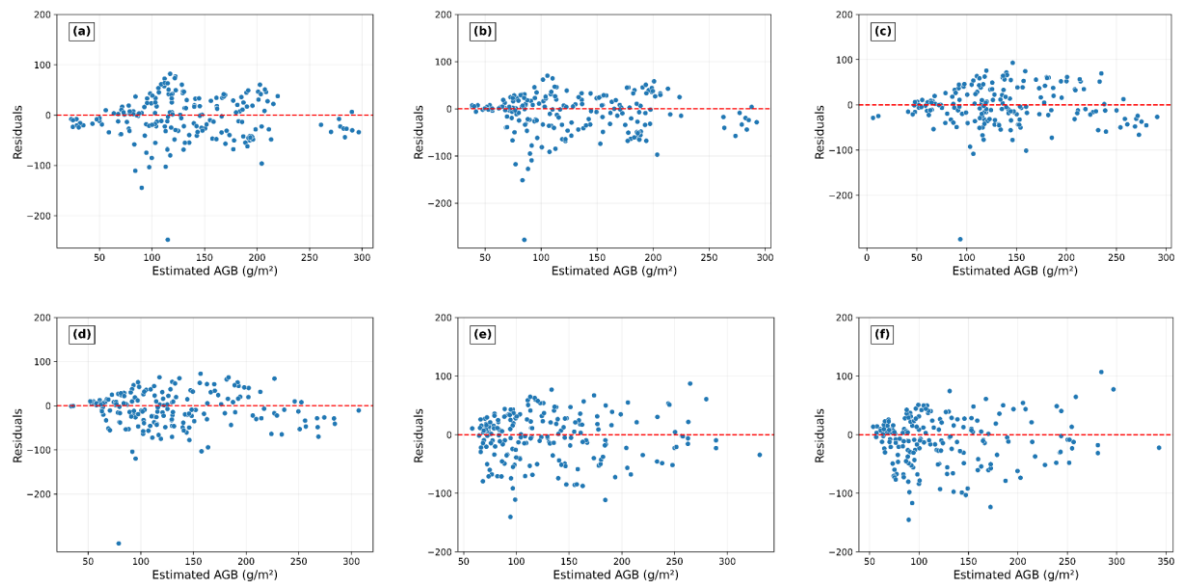


Figure 3.A1. Residual distribution of predicted versus ground-estimated AGB for six models (a-f), corresponding to the linear-PlanetScope, exponential-PlanetScope, linear-Sentinel-2, exponential-Sentinel-2, linear-Landsat 8-9, and exponential-Landsat 8-9

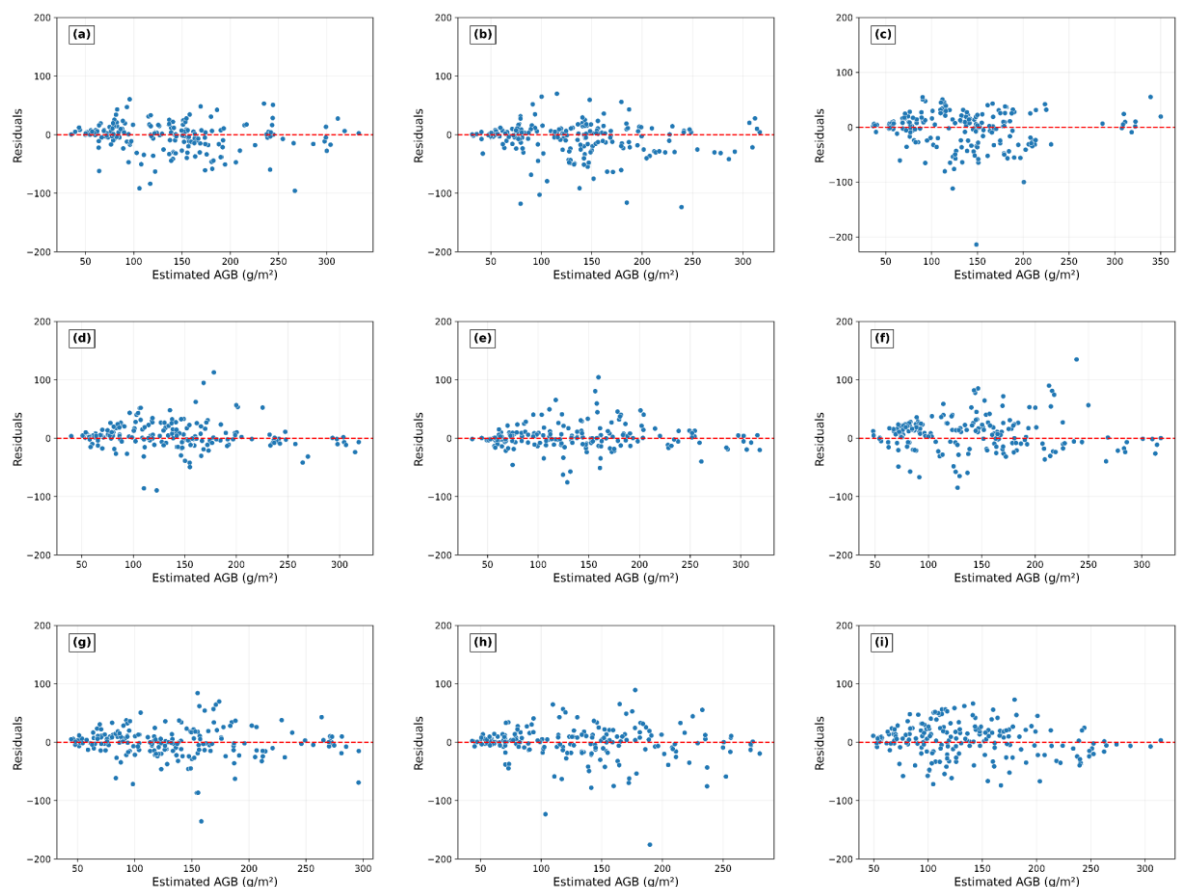


Figure 3.A2. Residual distribution of predicted versus ground-estimated AGB for six models (a-i), corresponding to the RF-PlanetScope, SVR-PlanetScope, ANN-PlanetScope, RF-Sentinel-2, SVR-Sentinel-2, ANN-Sentinel-2, RF-Landsat 8-9, SVR-Landsat 8-9, and ANN-Landsat 8-9

3.8. References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Abdollahi, A., Liu, Y., Pradhan, B., Huete, A., Dikshit, A., & Nguyen Tran, N. (2022). Short-time-series grassland mapping using Sentinel-2 imagery and deep learning-based architecture. *The Egyptian Journal of Remote Sensing and Space Science*, 25(3), 673–685. <https://doi.org/10.1016/j.ejrs.2022.06.002>
- Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., & Green, S. (2016). Satellite remote sensing of grasslands: From observation to management. *Journal of Plant Ecology*, 9(6), 649–671. <https://doi.org/10.1093/jpe/rtw005>
- Ali, I., Cawkwell, F., Dwyer, E., & Green, S. (2017). Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data—A Machine Learning Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(7), 3254–3264. <https://doi.org/10.1109/JSTARS.2016.2561618>
- Badreldin, N., Prieto, B., & Fisher, R. (2021). Mapping Grasslands in Mixed Grassland Ecoregion of Saskatchewan Using Big Remote Sensing Data and Machine Learning. *Remote Sensing*, 13(24), Article 24. <https://doi.org/10.3390/rs13244972>
- Barrett, B., Nitze, I., Green, S., & Cawkwell, F. (2014). Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. *Remote Sensing of Environment*, 152, 109–124. <https://doi.org/10.1016/j.rse.2014.05.018>
- Bates, D. M., & Watts, D. G. (1988). *Nonlinear regression analysis and its applications* (Vol. 2). Wiley New York.
- Bazzo, C. O. G., Kamali, B., Hütt, C., Bareth, G., & Gaiser, T. (2023). A Review of Estimation Methods for Aboveground Biomass in Grasslands Using UAV. *Remote Sensing*, 15(3), Article 3. <https://doi.org/10.3390/rs15030639>
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1), 281–305.
- Bishop, C. M. (2007). Pattern recognition and machine learning (information science and statistics). Springer New York.
- Braun, A., Wagner, J., & Hochschild, V. (2018). Above-ground biomass estimates based on active and passive microwave sensor imagery in low-biomass savanna ecosystems. *Journal of Applied Remote Sensing*, 12(4), 046027. <https://doi.org/10.1117/1.JRS.12.046027>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brereton, R. G., & Lloyd, G. R. (2010). Support Vector Machines for classification and regression. *Analyst*, 135(2), 230–267. <https://doi.org/10.1039/B918972F>

- Brownlee, J. (2016, November 6). How to Code a Neural Network with Backpropagation In Python (from scratch). *MachineLearningMastery.Com*. <https://www.machinelearningmastery.com/implement-backpropagation-algorithm-scratch-python/>
- Burger, R., Aouizerats, B., den Besten, N., Guillevic, P., Catarino, F., van der Horst, T., Jackson, D., Koopmans, R., Ridderikhoff, M., Robson, G., Zajdband, A., & de Jeu, R. (2024). The Biomass Proxy: Unlocking Global Agricultural Monitoring through Fusion of Sentinel-1 and Sentinel-2. *Remote Sensing*, 16(5), Article 5. <https://doi.org/10.3390/rs16050835>
- Chen, X., Zhu, X., Ye, X., Ouyang, G., Ma, L., & Wang, Q. (2024). Grassland Aboveground Biomass Estimation using Vegetation Indices with Machine Learning Algorithms. *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 10504–10507. <https://doi.org/10.1109/IGARSS53475.2024.10640949>
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15–18.
- Crippen, R., Buckley, S., Agram, P., Belz, E., Gurrola, E., Hensley, S., Kobrick, M., Lavalley, M., Martin, J., Neumann, M., Nguyen, Q., Rosen, P., Shimada, J., Simard, M., & Tung, W. (2016). NASADEM GLOBAL ELEVATION MODEL: METHODS AND PROGRESS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B4*, 125–128. XXIII ISPRS Congress, Commission IV (Volume XLI-B4) - 12–19 July 2016, Prague, Czech Republic. <https://doi.org/10.5194/isprs-archives-XLI-B4-125-2016>
- Cui, B., Huang, W., Ye, H., & Chen, Q. (2022). The Suitability of PlanetScope Imagery for Mapping Rubber Plantations. *Remote Sensing*, 14(5), Article 5. <https://doi.org/10.3390/rs14051061>
- Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests for Classification in Ecology. *Ecology*, 88(11), 2783–2792. <https://doi.org/10.1890/07-0539.1>
- De Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P. R., & Rowlings, D. W. (2021). Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. *Computers and Electronics in Agriculture*, 180, 105880. <https://doi.org/10.1016/j.compag.2020.105880>
- Defalque, G., Santos, R., Bungenstab, D., Echeverria, D., Dias, A., & Defalque, C. (2024). Machine learning models for dry matter and biomass estimates on cattle grazing systems. *Computers and Electronics in Agriculture*, 216, 108520. <https://doi.org/10.1016/j.compag.2023.108520>
- Dos Reis, A. A., Werner, J. P. S., Silva, B. C., Figueiredo, G. K. D. A., Antunes, J. F. G., Esquerdo, J. C. D. M., Coutinho, A. C., Lamparelli, R. A. C., Rocha, J. V., & Magalhães, P. S. G. (2020). Monitoring Pasture Aboveground Biomass and Canopy Height in an Integrated Crop–Livestock System Using Textural Information from PlanetScope Imagery. *Remote Sensing*, 12(16), Article 16. <https://doi.org/10.3390/rs12162534>

- Dusseux, P., Corpetti, T., Hubert-Moy, L., & Corgne, S. (2014). Combined use of multi-temporal optical and radar satellite images for grassland monitoring. *Remote Sensing*, 6(7), 6163–6182.
- Frazier, A. E., & Hemingway, B. L. (2021). A Technical Review of Planet Smallsat Data: Practical Considerations for Processing and Using PlanetScope Imagery. *Remote Sensing*, 13(19), Article 19. <https://doi.org/10.3390/rs13193930>
- Gao, T., Xu, Bin, Yang, Xiuchun, Jin, Yunxiang, Ma, Hailong, Li, Jinya, & Yu, H. (2013). Using MODIS time series data to estimate aboveground biomass and its spatio-temporal variation in Inner Mongolia's grassland between 2001 and 2011. *International Journal of Remote Sensing*, 34(21), 7796–7810. <https://doi.org/10.1080/01431161.2013.823000>
- Garrouste, E. L., Hansen, A. J., & Lawrence, R. L. (2016). Using NDVI and EVI to Map Spatiotemporal Variation in the Biomass and Quality of Forage for Migratory Elk in the Greater Yellowstone Ecosystem. *Remote Sensing*, 8(5), Article 5. <https://doi.org/10.3390/rs8050404>
- Gašparović, M., Medak, D., Pilaš, I., Jurjević, L., & Balenović, I. (2018). FUSION OF SENTINEL-2 AND PLANETSCOPE IMAGERY FOR VEGETATION DETECTION AND MONITORING. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1*, 155–160. ISPRS TC I Mid-term Symposium <q>Innovative Sensing – From Sensors to Methods and Applications</q> (Volume XLII-1) - 10–12 October 2018, Karlsruhe, Germany. <https://doi.org/10.5194/isprs-archives-XLII-1-155-2018>
- Ge, J., Meng, B., Liang, T., Feng, Q., Gao, J., Yang, S., Huang, X., & Xie, H. (2018). Modeling alpine grassland cover based on MODIS data and support vector machine regression in the headwater region of the Huanghe River, China. *Remote Sensing of Environment*, 218, 162–173. <https://doi.org/10.1016/j.rse.2018.09.019>
- Government of Canada, P. S. and P. C. (2002, July 1). *Prairies Ecozone+ evidence for key findings summary: Canadian biodiversity - ecosystem status and trends 2010. Evidence for key findings summary report No. 4 / published by the Canadian Councils of Resource Ministers. : En14-43/0-4-2014E-PDF - Government of Canada Publications - Canada.ca*. <https://publications.gc.ca/site/eng/9.629546/publication.html>
- Guerini Filho, M., Kuplich, T. M., & Quadros, F. L. F. D. (2020). Estimating natural grassland biomass by vegetation indices using Sentinel 2 remote sensing data. *International Journal of Remote Sensing*, 41(8), 2861–2876. <https://doi.org/10.1080/01431161.2019.1697004>
- Guillevic, P. C., Aouizerats, B., Burger, R., Den Besten, N., Jackson, D., Ridderikhoff, M., Zajdband, A., Houborg, R., Franz, T. E., Robertson, G. P., & De Jeu, R. (2024). Planet's Biomass Proxy for monitoring aboveground agricultural biomass and estimating crop yield. *Field Crops Research*, 316, 109511. <https://doi.org/10.1016/j.fcr.2024.109511>
- Gurney, R. J., Foster, J. L., & Parkinson, C. L. (1993). *Atlas of satellite observations related to global change*. Cambridge University Press.

- Harod, R. (2021, August 18). *Working with QA Bands and Bitmasks in Google Earth Engine*. Spatial Thoughts. <https://spatialthoughts.com/2021/08/19/qa-bands-bitmasks-gee/>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Haykin, S. (1994). *Neural networks: A comprehensive foundation*. Prentice Hall PTR.
- Hubert-Moy, L., Thibault, J., Fabre, E., Roza, C., Arvor, D., Corpetti, T., & Rapinel, S. (2019). Mapping Grassland Frequency Using Decadal MODIS 250 m Time-Series: Towards a National Inventory of Semi-Natural Grasslands. *Remote Sensing*, 11(24), Article 24. <https://doi.org/10.3390/rs11243041>
- Jargalsaikhan, M.-E., Ichikawa, D., Nagai, M., Indree, T., Katiyar, V., Munkhtur, D., & Dashdondog, E. (2024). Aboveground Biomass Estimation and Time Series Analyses in Mongolian Grasslands Utilizing PlanetScope Imagery. *Remote Sensing*, 16(5), Article 5. <https://doi.org/10.3390/rs16050869>
- Jargalsaikhan, M.-E., Nagai, M., Tumendemberel, B., Dashdondog, E., Katiyar, V., & Ichikawa, D. (2025). Adapting the High-Resolution PlanetScope Biomass Model to Low-Resolution VIIRS Imagery Using Spectral Harmonization: A Case of Grassland Monitoring in Mongolia. *Remote Sensing*, 17(8), Article 8. <https://doi.org/10.3390/rs17081428>
- Jobbágy, E. G., Sala, O. E., & Paruelo, J. M. (2002). Patterns and Controls of Primary Production in the Patagonian Steppe: A Remote Sensing Approach. *Ecology*, 83(2), 307–319. <https://doi.org/10.2307/2680015>
- Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., & Mattsson, E. (2015). Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest. *Remote Sensing*, 7(8), Article 8. <https://doi.org/10.3390/rs70810017>
- Khabbazan, S., Vermunt, P., Steele-Dunne, S., Ratering Arntz, L., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., & van der Sande, C. (2019). Crop Monitoring Using Sentinel-1 Data: A Case Study from The Netherlands. *Remote Sensing*, 11(16), Article 16. <https://doi.org/10.3390/rs11161887>
- Krieger, G., Moreira, A., Fiedler, H., Hajnsek, I., Werner, M., Younis, M., & Zink, M. (2007). TanDEM-X: A Satellite Formation for High-Resolution SAR Interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11), 3317–3341. <https://doi.org/10.1109/TGRS.2007.900693>
- Landsat Algorithms | Google Earth Engine*. (2024, October 27). Google for Developers. <https://developers.google.com/earth-engine/guides/landsat>
- Lee, S. (2025). *10 Practical Tips: Leveraging Cook's Distance for Outlier Detection*. <https://www.numberanalytics.com/blog/10-tips-leveraging-cooks-distance-outlier-detection>

- Li, C., Zhou, L., & Xu, W. (2021). Estimating Aboveground Biomass Using Sentinel-2 MSI Data and Ensemble Algorithms for Grassland in the Shengjin Lake Wetland, China. *Remote Sensing*, 13(8), Article 8. <https://doi.org/10.3390/rs13081595>
- Li, F., Jiang, L., Wang, X., Zhang, X., Zheng, J., & Zhao, Q. (2013). Estimating grassland aboveground biomass using multitemporal MODIS data in the West Songnen Plain, China. *Journal of Applied Remote Sensing*, 7(1), 073546. <https://doi.org/10.1117/1.JRS.7.073546>
- Li, H., Li, F., Xiao, J., Chen, J., Lin, K., Bao, G., Liu, A., & Wei, G. (2024). A machine learning scheme for estimating fine-resolution grassland aboveground biomass over China with Sentinel-1/2 satellite images. *Remote Sensing of Environment*, 311, 114317. <https://doi.org/10.1016/j.rse.2024.114317>
- Li, J., Ravi, S., Wang, G., Van Pelt, R. S., Gill, T. E., & Sankey, J. B. (2022). Woody plant encroachment of grassland and the reversibility of shrub dominance: Erosion, fire, and feedback processes. *Ecosphere*, 13(3), e3949. <https://doi.org/10.1002/ecs2.3949>
- Liang, S., & Wang, J. (Eds.). (2020). Chapter 1—A systematic view of remote sensing. In *Advanced Remote Sensing (Second Edition)* (pp. 1–57). Academic Press. <https://doi.org/10.1016/B978-0-12-815826-5.00001-5>
- Liu, S., Cheng, F., Dong, S., Zhao, H., Hou, X., & Wu, X. (2017). Spatiotemporal dynamics of grassland aboveground biomass on the Qinghai-Tibet Plateau based on validated MODIS NDVI. *Scientific Reports*, 7(1), 4182. <https://doi.org/10.1038/s41598-017-04038-4>
- Liu, W., Xu, C., Zhang, Z., De Boeck, H., Wang, Y., Zhang, L., Xu, X., Zhang, C., Chen, G., & Xu, C. (2023). Machine learning-based grassland aboveground biomass estimation and its response to climate variation in Southwest China. *Frontiers in Ecology and Evolution*, 11, 1146850. <https://doi.org/10.3389/fevo.2023.1146850>
- Malenovský, Z., Homolová, L., Zurita-Milla, R., Lukeš, P., Kaplan, V., Hanuš, J., Gastellu-Etchegorry, J.-P., & Schaepman, M. E. (2013). Retrieval of spruce leaf chlorophyll content from airborne image data using continuum removal and radiative transfer. *Remote Sensing of Environment*, 131, 85–102. <https://doi.org/10.1016/j.rse.2012.12.015>
- Manfreda, S., McCabe, M. F., Miller, P. E., Lucas, R., Pajuelo Madrigal, V., Mallinis, G., Ben Dor, E., Helman, D., Estes, L., Ciraolo, G., Müllerová, J., Tauro, F., De Lima, M. I., De Lima, J. L. M. P., Maltese, A., Frances, F., Caylor, K., Kohv, M., Perks, M., ... Toth, B. (2018). On the Use of Unmanned Aerial Systems for Environmental Monitoring. *Remote Sensing*, 10(4), Article 4. <https://doi.org/10.3390/rs10040641>
- Markham, B., Jenstrom, D., Sauer, B., Pszcolka, S., Dulski, V., Hair, J., McCorkel, J., Kvaran, G., Thome, K., Montanaro, M., Pedelty, J., Anderson, C., Choate, M., Barsi, J., Kaita, E., & Miller, J. (2020). Landsat 9 mission update and status. *Earth Observing Systems XXV*, 11501, 96–102. <https://doi.org/10.1117/12.2569748>

- McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). *Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning*. <https://doi.org/10.1175/BAMS-D-18-0195.1>
- MODIS Web. (n.d.). Retrieved April 18, 2025, from <https://modis.gsfc.nasa.gov/data/dataproduct/>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- Morais, T. G., Teixeira, R. F. M., Figueiredo, M., & Domingos, T. (2021). The use of machine learning methods to estimate aboveground biomass of grasslands: A review. *Ecological Indicators*, 130, 108081. <https://doi.org/10.1016/j.ecolind.2021.108081>
- Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F. A., Schwarz, L.-M., Ghazaryan, G., Schultz, J., Malenovský, Z., & Dubovyk, O. (2022). Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. *Remote Sensing of Environment*, 282, 113262. <https://doi.org/10.1016/j.rse.2022.113262>
- Naidoo, L., van Deventer, H., Ramoelo, A., Mathieu, R., Nondlazi, B., & Gangat, R. (2019). Estimating above ground biomass as an indicator of carbon storage in vegetated wetlands of the grassland biome of South Africa. *International Journal of Applied Earth Observation and Geoinformation*, 78, 118–129. <https://doi.org/10.1016/j.jag.2019.01.021>
- Neter, J., Wasserman, W., & Kutner, M. H. (1983). *Applied linear regression models*. Richard D. Irwin.
- Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: A primer for ecologists. *The Quarterly Review of Biology*, 83(2), 171–193. <https://doi.org/10.1086/587826>
- Paltsyn, M. Yu., Gibbs, J. P., Iegorova, L. V., & Mountrakis, G. (2017). Estimation and Prediction of Grassland Cover in Western Mongolia Using MODIS-Derived Vegetation Indices. *Rangeland Ecology & Management*, 70(6), 723–729. <https://doi.org/10.1016/j.rama.2017.05.005>
- Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Parracciani, C., Gigante, D., Bonini, F., Grassi, A., Morbidini, L., Pauselli, M., Valenti, B., Lilli, E., Antonielli, F., & Vizzari, M. (2024). Leveraging Google Earth Engine for a More Effective Grassland Management: A Decision Support Application Perspective. *Sensors*, 24(3), Article 3. <https://doi.org/10.3390/s24030834>
- Pasquarella, V. (2023, October 30). All Clear with Cloud Score+. *Medium*. <https://medium.com/google-earth/all-clear-with-cloud-score-bd6ee2e2235e>
- Peeters, A. (2009). Importance, evolution, environmental impact and future challenges of grasslands and grassland-based systems in Europe. *Grassland Science*, 55(3), 113–125. <https://doi.org/10.1111/j.1744-697X.2009.00154.x>

- Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sensing*, 12(14), Article 14. <https://doi.org/10.3390/rs12142291>
- PlanetScope | Planet Documentation. (2025, May 5). <https://docs.planet.com/data/imagery/planetscope/>
- Qayyum, N., Ghuffar, S., Ahmad, H. M., Yousaf, A., & Shahid, I. (2020). Glacial Lakes Mapping Using Multi Satellite PlanetScope Imagery and Deep Learning. *ISPRS International Journal of Geo-Information*, 9(10), Article 10. <https://doi.org/10.3390/ijgi9100560>
- Qin, Q., Xu, D., Hou, L., Shen, B., & Xin, X. (2021). Comparing vegetation indices from Sentinel-2 and Landsat 8 under different vegetation gradients based on a controlled grazing experiment. *Ecological Indicators*, 133, 108363. <https://doi.org/10.1016/j.ecolind.2021.108363>
- Rankin, J., & Komainda, M. (2023). *Practical use of the rising plate meter for yield estimation*.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- Reinermann, S., Asam, S., & Kuenzer, C. (2020). Remote Sensing of Grassland Production and Management—A Review. *Remote Sensing*, 12(12), Article 12. <https://doi.org/10.3390/rs12121949>
- Riley, J. L., Green, S. E., & Brodribb, K. E. (2007). *A conservation blueprint for Canada's prairies and parklands*. Nature Conservancy of Canada.
- Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. *Journal of Computer-Aided Molecular Design*, 36(5), 355–362. <https://doi.org/10.1007/s10822-022-00442-9>
- Rosenqvist, A., Shimada, M., Suzuki, S., Ohgushi, F., Tadono, T., Watanabe, M., Tsuzuku, K., Watanabe, T., Kamijo, S., & Aoki, E. (2014). Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. *Remote Sensing of Environment*, 155, 3–12. <https://doi.org/10.1016/j.rse.2014.04.011>
- Roy, D. P., Wulder, M. A., Loveland, T. R., C.e., W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R., Sheng, Y., Vermote, E. F., Belward, A. S., Bindaschadler, R., Cohen, W. B., Gao, F., ... Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, 145, 154–172. <https://doi.org/10.1016/j.rse.2014.02.001>
- Schmidt, M., Carter, J., Stone, G., & O'Reagain, P. (2016). Integration of Optical and X-Band Radar Data for Pasture Biomass Estimation in an Open Savannah Woodland. *Remote Sensing*, 8(12), Article 12. <https://doi.org/10.3390/rs8120989>

- Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., & Hostert, P. (2022). Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. *Remote Sensing of Environment*, 269, 112795. <https://doi.org/10.1016/j.rse.2021.112795>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1), 60. <https://doi.org/10.1186/s40537-019-0197-0>
- Shorthouse, J. D. (2010). Ecoregions of Canada's prairie grasslands. *Arthropods of Canadian Grasslands*, 1, 53–81.
- Smith, R. E., Veldhuis, H., Mills, G. F., Eilers, R. G., Fraser, W. R., & Lelyk, G. W. (1998). *Terrestrial Ecozones, Ecoregions and Ecodistricts of Manitoba*.
- Staples, G. C. (2005). RADARSAT-2. In M. L. Schwartz (Ed.), *Encyclopedia of Coastal Science* (pp. 784–785). Springer Netherlands. https://doi.org/10.1007/1-4020-3880-1_252
- Sullivan, J. H., Warkentin, M., & Wallace, L. (2021). So many ways for assessing outliers: What really works and does it matter? *Journal of Business Research*, 132, 530–543. <https://doi.org/10.1016/j.jbusres.2021.03.066>
- Talebizadeh, M., & Moridnejad, A. (2011). Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models. *Expert Systems with Applications*, 38(4), 4126–4135. <https://doi.org/10.1016/j.eswa.2010.09.075>
- Thornton, M. M., Shrestha, R., Wei, Y., Thornton, P. E., & Kao, S.-C. (2020). *Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4*. ORNL Distributed Active Archive Center. <https://doi.org/10.3334/ORNLDAAC/1840>
- Tian, L., Wu, X., Tao, Y., Li, M., Qian, C., Liao, L., & Fu, W. (2023). Review of Remote Sensing-Based Methods for Forest Aboveground Biomass Estimation: Progress, Challenges, and Prospects. *Forests*, 14(6), Article 6. <https://doi.org/10.3390/f14061086>
- Tian, Y., & Zhang, Y. (2022). A comprehensive survey on regularization strategies in machine learning. *Information Fusion*, 80, 146–166. <https://doi.org/10.1016/j.inffus.2021.11.005>
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., ... Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–24. <https://doi.org/10.1016/j.rse.2011.05.028>
- Uyeda, K. A., Stow, Douglas A., Roberts, Dar A., & Riggan, P. J. (2017). Combining ground-based measurements and MODIS-based spectral vegetation indices to track biomass accumulation in post-fire chaparral. *International Journal of Remote Sensing*, 38(3), 728–741. <https://doi.org/10.1080/01431161.2016.1271477>
- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science & business media.
- Vawda, M. I., Lottering, R., Mutanga, O., Peerbhay, K., & Sibanda, M. (2024). Comparing the Utility of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN)

- on Sentinel-2 MSI to Estimate Dry Season Aboveground Grass Biomass. *Sustainability*, 16(3), Article 3. <https://doi.org/10.3390/su16031051>
- Vreugdenhil, M., Wagner, W., Bauer-Marschallinger, B., Pfeil, I., Teubner, I., Rüdiger, C., & Strauss, P. (2018). Sensitivity of Sentinel-1 Backscatter to Vegetation Dynamics: An Austrian Case Study. *Remote Sensing*, 10(9), Article 9. <https://doi.org/10.3390/rs10091396>
- Wang, G., Liu, Shuman, Liu, Tingxi, Fu, Zhiyuan, Yu, Jingshan, & Xue, B. (2019). Modelling above-ground biomass based on vegetation indexes: A modified approach for biomass estimation in semi-arid grasslands. *International Journal of Remote Sensing*, 40(10), 3835–3854. <https://doi.org/10.1080/01431161.2018.1553319>
- Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R. B., & Chang, Q. (2019). Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154, 189–201. <https://doi.org/10.1016/j.isprsjprs.2019.06.007>
- Wang, L., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal*, 4(3), 212–219. <https://doi.org/10.1016/j.cj.2016.01.008>
- Wang, T., Zhou, W., Xiao, J., & Xie, L. (2023). Estimating the grassland aboveground biomass based on remote sensing data and machine learning algorithm. *Journal of Glaciology and Geocryology*, 45(2), Article 2. <https://doi.org/10.7522/j.issn.1000-0240.2023.0057>
- Wang, Z., Ma, Y., Zhang, Y., & Shang, J. (2022). Review of Remote Sensing Applications in Grassland Monitoring. *Remote Sensing*, 14(12), Article 12. <https://doi.org/10.3390/rs14122903>
- Wu, W., Sun, R., Liu, L., Liu, X., Yu, H., Ma, Q., Qi, M., Li, L., Li, Y., Zhou, G., & Xu, Z. (2023). Precipitation consistently promotes, but temperature inversely drives, biomass production in temperate vs. alpine grasslands. *Agricultural and Forest Meteorology*, 329, 109277. <https://doi.org/10.1016/j.agrformet.2022.109277>
- Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., ... Cook, B. D. (2022). Fifty years of Landsat science and impacts. *Remote Sensing of Environment*, 280, 113195. <https://doi.org/10.1016/j.rse.2022.113195>
- Xu, D., & Guo, X. (2015). Some Insights on Grassland Health Assessment Based on Remote Sensing. *Sensors*, 15(2), Article 2. <https://doi.org/10.3390/s150203070>
- Xu, P., Ji, X., Li, M., & Lu, W. (2023). Small data machine learning in materials science. *Npj Computational Materials*, 9(1), 1–15. <https://doi.org/10.1038/s41524-023-01000-z>
- YANG, Y.-H., & PIAO, S.-L. (2006). Variations in grassland vegetation cover in relation to climatic factors on the Tibetan Plateau. *Chinese Journal of Plant Ecology*, 30(1), 1.

- Yao, Y., & Ren, H. (2024). Estimation of grassland aboveground biomass in northern China based on topography-climate-remote sensing data. *Ecological Indicators*, 165, 112230. <https://doi.org/10.1016/j.ecolind.2024.112230>
- Yuan, Z., Wang, L.-N., & Ji, X. (2014). Prediction of concrete compressive strength: Research on hybrid models genetic based algorithms and ANFIS. *Advances in Engineering Software*, 67, 156–163. <https://doi.org/10.1016/j.advengsoft.2013.09.004>
- Zeng, N., Ren, X., He, H., Zhang, L., Li, P., & Niu, Z. (2021). Estimating the grassland aboveground biomass in the Three-River Headwater Region of China using machine learning and Bayesian model averaging. *Environmental Research Letters*, 16(11), 114020. <https://doi.org/10.1088/1748-9326/ac2e85>
- Zeng, N., Ren, X., He, H., Zhang, L., Zhao, D., Ge, R., Li, P., & Niu, Z. (2019). Estimating grassland aboveground biomass on the Tibetan Plateau using a random forest algorithm. *Ecological Indicators*, 102, 479–487. <https://doi.org/10.1016/j.ecolind.2019.02.023>
- Zhang, M., Sun, J., Wang, Y., Li, Y., & Duo, J. (2025). State-of-the-art and challenges in global grassland degradation studies. *Geography and Sustainability*, 6(2), 100229. <https://doi.org/10.1016/j.geosus.2024.08.008>
- Zhao, F., Xu, B., Yang, X., Jin, Y., Li, J., Xia, L., Chen, S., & Ma, H. (2014). Remote Sensing Estimates of Grassland Aboveground Biomass Based on MODIS Net Primary Productivity (NPP): A Case Study in the Xilingol Grassland of Northern China. *Remote Sensing*, 6(6), Article 6. <https://doi.org/10.3390/rs6065368>
- Zhou, L., Kaufmann, R., Tian, 1 Y, Myneni, R., & Tucker, C. (2003). Relation between interannual variations in satellite measures of northern forest greenness and climate between 1982 and 1999. *Journal of Geophysical Research: Atmospheres*, 108(D1), ACL-3.
- Zhao, S., Zhu, X., Tan, X., & Tian, J. (2025). Spectrotemporal fusion: Generation of frequent hyperspectral satellite imagery. *Remote Sensing of Environment*, 319, 114639. <https://doi.org/10.1016/j.rse.2025.114639>
- Zhou, W., Li, H., Xie, L., Nie, X., Wang, Z., Du, Z., & Yue, T. (2021). Remote sensing inversion of grassland aboveground biomass based on high accuracy surface modeling. *Ecological Indicators*, 121, 107215. <https://doi.org/10.1016/j.ecolind.2020.107215>
- Zocca, V., Spacagna, G., Slater, D., & Roelants, P. (2017). *Python Deep Learning*. Packt Publishing Ltd.

4. OVERALL SYNTHESIS

4.1. Study Findings and Implications

Research on grassland ecosystems, specifically monitoring grassland using remote sensing, is limited in the Canadian Prairies as a primary location of grassland species. Distribution of different grassland types, such as native, tame and mixed grasses, across Manitoba's Prairie regions are a dynamic phenomenon due to interior (e.g, soil, moisture, geography) and exterior (e.g., anthropogenic pressure and climate variability) factors that need to be evaluated timely for a large area. From economic and ecological aspects, it is essential to provide a reliable mapping approach to determine the area and location of tame and native grasses (Pinna et al., 2024).

The analysis carried out in Chapter 2 indicated that vegetative indices (VIs) obtained from satellite imagery and topographic variables (e.g., elevation, slope and aspect) can effectively distinguish the three primary categories of grassland. The study developed a new AI-based hierarchical strategy to improve machine learning classification robustness and to prevent other land cover classes from being included in the final grasslands mapping, which can potentially generate errors. Multi-source remote sensing data, including microwave (Sentinel-1 sensor) and optical (Sentinel-2 and Landsat 8 and 9 sensors) images, were tested, and the best combination was Sentinel-1+Sentinel-2 in land use and grassland classification. To include phenological growth information in the models, multi-temporal data were collected across the Prairie area (around 69,000 km²) during the growing season using the Google Earth Engine (GEE) platform. Multi-temporal compositions were used to effectively distinguish between different types of grassland. Three well-known machine learning methods, like random forest (RF), support vector machine (SVM), and gradient tree boosting (GTB) were implemented on GEE to map grassland types (e.i., native, tame and mixed) in three ecoregions, including the Aspen Parkland, the Lake Manitoba Plain and the Southwest Manitoba Uplands. The RF model achieved the highest classification performance, with an overall accuracy (OA) of 72.62% and a kappa value of 0.58. GTB performed slightly lower than RF, while SVM was weaker in comparison. A part of the classification accuracy improvement came from feature selection done using analysis of variance (ANOVA), which identified significant features by comparing the means of different groups. The listing of features according to their scores from higher to lower also showed that optical features (like spectral

bands and VIs) outperformed SAR features (like VV and VH) in all classifier models. Around 28% of the total Prairie region in Manitoba consisted of grasslands, with a combination of 53.18% tame, 25.69% native, and 21.13% mixed grasses. The study also revealed that the Lake Manitoba ecoregion contains the greatest amount of native grasses, primarily concentrated around Lake Manitoba. The application of the fusion technique was also assessed in mapping grasslands when it was not feasible to use multi-temporal data. Finally, a web map was created to show grassland results within the Soil Landscapes of Canada (SLC) polygons, relating soil landscapes to grassland distribution and providing valuable information for decision-makers and researchers.

In Chapter 3, the research aimed to analyze one of the most important and challenging biophysical parameters of grassland, namely aboveground biomass (AGB), using the most available remote sensing data and various AI models. AGB is a key indicator of vegetation health, productivity, and carbon storage (Asam et al., 2015). There are only a few studies that used remote sensing data to estimate Canadian grassland AGB. This study aimed to propose a comprehensive methodology for biomass estimation in the Manitoba Prairie region using different sensor types and models. Because of the limited ground truth data, Planet's Biomass Proxy product, which has daily temporal resolution, was used to generate synthetic biomass samples using an exponential regression model across the growing season of the years 2021-2023. Three remote sensing data, such as PlanetScope, Sentinel-2 and Landsat 8 and 9 (Landsat 8-9), were utilized in biomass estimation models, like parametric (linear and exponential regression) and machine learning models [RF, support vector regression (SVR) and artificial neural network (ANN) regression]. The results indicated that machine learning approaches significantly outperformed traditional parametric models, particularly when VIs were combined with climatic (precipitation and temperature) and topographic [digital elevation models (DEM)] predictors. SVR and RF models achieved the highest predictive accuracies, with R^2 values up to 0.90. The inclusion of environmental variables enhanced model performance significantly, with R^2 improvements ranging from 0.11 to 0.35 across various models and sensors. Dimensionality reduction using principal component analysis (PCA) further helped in handling multicollinearity, which can inflate R-squared values in regression models. The results indicated that RF models outperformed linear parametric models in addressing this issue. Cross-validation (k-fold) confirmed that high R^2 values in some

test-set splits could be due to favourable data partitioning, with cross-validation scores offering more reliable estimates of model generalizability. For instance, ANN models using Landsat data achieved the highest R^2 among the three sensors; however, their k-fold R^2 values indicated lower performance compared to the others.

The results highlight the strength of advanced machine learning models in biomass modelling in the presence of multicollinearity and data heterogeneity. Among different models, RF offered a robust and high-performing model for grassland biomass estimation, while among the different data sources, PlanetScope with 3 m spatial resolution produced the most detailed biomass maps. Although coarser-resolution satellite imagery such as Landsat 8–9 also achieved high performance, it was more suitable for larger farm-scale applications. The advantage of the integration of environmental variables demonstrates the need for climatic context in remote sensing-based models of productivity. The findings of the comparison of sensors imply that operational biomass monitoring programs can be customized according to the sensor availability, budget, and size of the project without losing much in terms of accuracy. Lastly, the validated framework highlights the necessity of cross-validation.

4.2. Study Recommendations

Although this study represents a pioneering effort in monitoring grassland distribution, area, and biomass in the Canadian Prairies of Manitoba using remote sensing approaches, several recommendations can help improve future research in remote sensing-based grassland evaluation. First, increasing the number of field samples at different phenological stages of grass and over multiple years would provide a more representative understanding of grassland conditions. A larger sample size enhances the statistical robustness of models and offers a more realistic depiction of on-ground variability. Second, incorporating additional influential variables, such as soil type, more detailed meteorological parameters, and biophysical indicators like leaf area index (LAI) and canopy height, could significantly improve model accuracy.

Unmanned aerial vehicles (UAVs) or drones are also suggested to measure and estimate the biophysical parameters (e.g., canopy height and vegetation cover) or provide high-resolution ground truth information at the field scale (Dusseux et al., 2015; Schellberg

et al., 2008). Observations with the use of UAVs may be an effective way to fill in the gap between the satellite-based and field-based measurements, providing more spatial detail and more accurate biomass estimation models. In the future, it may be of interest to investigate the application of advanced deep learning models, which have been extremely promising in identifying complex spatial and temporal relationships in remote sensing data, including Convolutional Neural Networks (CNNs) (Kattenborn et al., 2021). These models have the potential to surpass the traditional machine learning algorithms in that they automatically.

The combination of Synthetic Aperture Radar (SAR) and optical remote sensing data in biomass estimation should be investigated in the future. Different multi-polarization data can be explored for biomass estimation. The individual sensor types have their own advantages. SAR is good at giving structural information and is not affected by cloud cover, whereas optical data is good at capturing spectral vegetation properties. The complementary nature of these data sources can be taken advantage of by using the combination of these data sources to overcome the weaknesses of each data source separately (Koch, 2010). With ongoing developments in sensor technologies, multisensor fusion will probably serve as the key element in operational biomass monitoring systems.

The research datasets were not evenly collected across the growing season or over multiple years for all fields where biomass samples were taken. Therefore, it is advised to conduct field campaigns across multiple time points within the growing season, which would enable capture of seasonal variability and the detection of abrupt changes (e.g., stress or disturbance events) across different landscapes and grassland types.

4.3. References

- Asam, S., Klein, D., & Dech, S. (2015). Estimation of grassland use intensities based on high spatial resolution LAI time series. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7-W3, 285–291. 36th International Symposium on Remote Sensing of Environment (Volume XL-7/W3) - 11–15 May 2015, Berlin, Germany. <https://doi.org/10.5194/isprsarchives-XL-7-W3-285-2015>
- Dusseux, P., Hubert-Moy, L., Corpetti, T., & Vertès, F. (2015). Evaluation of SPOT imagery for the estimation of grassland biomass. *International Journal of Applied Earth Observation and Geoinformation*, 38, 72–77. <https://doi.org/10.1016/j.jag.2014.12.003>

- Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 24–49. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>
- Koch, B. (2010). Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6), 581–590. <https://doi.org/10.1016/j.isprsjprs.2010.09.001>
- Pinna, D., Pezzuolo, A., Cogato, A., Pornaro, C., Macolino, S., & Marinello, F. (2024). Applications of satellite platforms and machine learning for mapping and monitoring grasslands and pastures: A systematic and comprehensive review. *Smart Agricultural Technology*, 9, 100571. <https://doi.org/10.1016/j.atech.2024.100571>
- Schellberg, J., Hill, M. J., Gerhards, R., Rothmund, M., & Braun, M. (2008). Precision agriculture on grassland: Applications, perspectives and constraints. *European Journal of Agronomy*, 29(2), 59–71. <https://doi.org/10.1016/j.eja.2008.05.005>