MONITORING AND PREDICTION OF PASTURE QUALITY AND PRODUCTIVITY USING PLANET SCOPE SATELLITE DATA FOR SUSTAINABLE LIVESTOCK PRODUCTION SYSTEMS IN COLOMBIA.

BY

ANUSHKA GHILDIYAL

2488539G

A dissertation submitted in partial fulfilment of the degree of Master of Science in Geoinformation Technology & Cartography, University of Glasgow

August 2020

Declaration

I, Anushka Ghildiyal declare that this dissertation is the product of my own work, except where indicated, and has not been submitted by myself or any other person for any degree at this or any other university.

Anushka Ghildiyal

2488539G

(31/08/2020)

Acknowledgement

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor, Dr. Brian Barrett, whose expertise was priceless in framing the research topic and methodology. It is whole-heartedly valued that your great advice for my study was monumental towards the success of this study.

I would also like to pay my special regards to Mr. Niantang Liu for the suggestions and all the help he provided in my supervisor's absence.

In addition, I would like to thank my parents for their sensible guidance. Finally, my friends who were of great support in reflecting over any issues faced, as well as for providing a much-needed form of escape from my studies.

List of Figures

List of Tables

Table of Contents

Abstract

As the population increases, demand for food increases too, which has led to large-scale land conversion to improve livestock production in Colombia. Fulfilling these criteria of increasing demand in a sustainable way is a challenge and remote sensing data provides an accurate method to support this task. In this study, Planet Scope multispectral satellite datasets and coincident field measurements acquired over test fields in the study area (Patía) of September 2018 was used. Fresh and dry weight biomass was calculated and forage quality analyses, crude protein (CP), in vitro dry matter digestibility (IVDMD), Ash and standing biomass dry weight (DM) was carried out in the forage nutritional quality laboratory of International Centre for Tropical Agriculture (CIAT). Field data was related to the remote sensing data using the random forest regression algorithm. R was required for the statistical analysis, to figure out the model performance for IVDMD, CP, Ash and DM. This project also investigated the spatial distribution of livestock which is affected by quality and area of potential forage zones. The R^2 values of the regression models were 0.74 for IVDMD, 0.69 for CP, 0.38 for Ash and 0.49 for DM using a predictor combination of vegetation indices, simple ratios and bands.

1. Introduction

Grasslands occupy ~40% of the world's land area, excluding Antarctica and Greenland (Suttie et al., 2005) and can be defined as land devoted to forage production for harvesting through grazing / browsing, cutting, or both, and used for certain agricultural purposes such as the production of renewable energy (Peeters et al., 2014). In their study Murray and Rohweder (2000) characterised grasslands as terrestrial habitats dominated by grass and shrub vegetation and sustained by burning, grazing, drought and/or freezing temperatures. From an agricultural point of view, grasslands are the cheapest feed supply for livestock (Suttie et al., 2005). Therefore, grassland management such as clearing trees and bushes and grazing or is necessary to maintain semi-natural grasslands rich in species (Tälle, 2018). Forage grasslands are used to provide food for livestock and are estimated to account for 26 % of the land area and 70 % of the agricultural area worldwide (Capstaff & Miller, 2018; FAO, 2010). Threats like agricultural intensification and urbanisation (Hooftman & Bullock, 2012; Kemp & Michalk, 2007) lead to land abandonment (Isselstein et al., 2005; Valkó et al., 2018) and climate change (Dangal et al., 2016; Lamarque et al., 2014; Waldén, 2018) which calls for the conservation of semi natural grasslands.

Herbivores influence the environment by grazing, which in turn influences the distribution of plant species (Palmer et al., 2005). Their activities are affected by the accessibility and quality of potential forage (Merkle et al., 2016; Palmer et al., 2005; Raab et al., 2020; Raynor et al., 2016) which means that it is necessary to get spatially detailed information about forage quantity and quality for the successful management of grasslands. As a lot of components work together to influence herbivore grazing, it is suggested that both forage quantity and quality should be investigated (Felton et al., 2018). Forage quality can be defined as the capacity of forage to fulfil the nutrient requirements of animals and it is determined by chemical and physical biomass characteristics (Collins et al., n.d.; Guo et al., 2010) or it can refer to how much animals consume a forage and how easily nutrients are processed into animal products in the forage (Fulgueira et al., 2007). Some useful parameters are dry matter (everything contained in a feed sample except water; this includes protein, fibre, fat, minerals, etc.), crude protein (the crude protein content of a feed sample represents the total nitrogen (N) in the diet), in vitro dry matter digestibility (IVDMD) method (extensively used to evaluate the nutritional value of ruminant feeds) and Ash (the total mineral content of a forage or diet) (Forage Quality Parameters Explained Agronomy Fact Sheet Series, 2016).

Traditional methods for measuring forage biomass depend on time-consuming hand cutting and drying of randomly selected samples across the area (Liu et al., 2019). Remote sensing approaches have considerable potential to track grassland and forage habitat characteristics, such as biomass, forage quality, and species identification and diversity (Wachendorf, 2018). Remotely sensed satellite imagery provides timely and repeatable spatial explicit information as compared to collecting field data manually (Ali et al. 2016). Ali et al., (2017); Boschetti et al., (2006); Zha et al., (2003) have demonstrated the feasibility of the grassland yield estimation using satellite remote sensing data using various statistical analysis methods. Researchers have developed and explored different regression models (e.g., linear, power, logarithmic, multiple linear) for grassland biomass estimation (Belgiu & Drăgu, 2016) as well as machine learning techniques like Random Forest (RF) (Ali et al., 2014), Artificial Neural Networks (ANN) (Ali et al., 2017) or Support Vector Machines (SVM) (Liakos et al., 2018). The classification and regression tree (CART) model, first introduced by (Breiman, 2001; Lima et al., 2015) is the most used decision tree. It uses a randomly selected subset of available predictors at each decision branch (bagging approach), each tree is grown over a bootstrap sample from the training dataset and the RF model 's output is the average output for all trees (Breiman, 2001; Lima et al., 2015). Random forests on the other hand is an ensemble learning technique, which has proved to be efficient for regression techniques as compared to ANN and SVM as it is not prone to overfitting, works well with missing data values and has low bias (Chan et al., 2012; Parente et al., 2019).

Using remote sensing to estimate the biophysical parameters of grasslands through vegetation indices and individual bands is shown in several studies (Loozen et al., 2019; Tong & He, 2017a). Both linear and nonlinear based machine learning approaches like lasso and ridge regression between predictor and biophysical variables have shown promising results (Mutanga et al., 2004; Zandler et al., 2015). The correlation between a predictor variable and the corresponding biophysical variable may change its slope, based on the phenological process, when time series data is applied; in such cases the RF regression algorithm can be superior to linear regression techniques (Beckschäfer et al., 2014; Raab et al., 2020; Strobl et al., 2007).

Hyperspectral sensors provide accurate spectral information. However, due to the expense and complexity of hyperspectral data, the reduction of the spectral data range and the recognition of the best spectral characteristics of hyperspectral information are still the most critical objectives which will promote easy field applications (Li et al., 2014; Reddersen et al., 2014). Spectral reflection measurements have been commonly used to classify grassland biomass obtained from handheld hyperspectral radiometers (Kawamura et al., 2011; Mutanga et al., 2004; Vescovo et al., 2012), but can contain significant quantities of redundant material (Reddersen et al., 2014; Wachendorf, 2018).

The term forage is defined as herbaceous plants or plant parts fed to domestic animals and it is divided into two major groups: grasses and legumes (Franklin & Martin, 1993). Hundreds of grasses are produced in tropical pastures and constitute a massive and economically substantial resource for tropics. Brachiaria grasses are the most widely cultivated forages in tropical South America, covering more than 80 million hectares (Boddey et al., 2004). The balanced mix of Brachiaria hybrids used in this study like Mulato II, Cayman yield superior results to those obtained with other Brachiaria genus grasses (Franklin & Martin, 1993). It provides excellent quality forage production capacity, with the best concentration of protein and high digestibility in Brachiaria genus pastures. It has easier management of high animal loads which significantly increases the production of meat and milk per hectare as well as increased drought tolerance and adaptation to acidic soils (Varieties - Tropical Seeds, 2020). Brachiaria decumbens and Brachiaria humidicola is an effective forage used in permanent pastures. It is high-yielding and forms low-leafy stands which function well on infertile soils (Signal Grass (Brachiaria Decumbens) | Feedipedia, 2020). This is palatable to all animal groups and can tolerate intense grazing (Loch, 1977). The grass species Megathyrsus maximus cv. Mombasa has one of the greatest known dry matter (DM) production potential in subtropical and tropical environments and can yield approximately 33 t/ha^{-1} annual dry matter production (Galindo et al., 2018). Dichanthium aristatum cv. Angleton is well eaten by all classes of stock when leafy, grows on poorly drained and seasonally flooded soils, withstands heavy grazing and suppresses invasive weeds such as *Phyla canescens* (Fact Sheet - Dichanthium Aristatum, 2020). The difference in the areas does not affect the cattle-rearing systems which are based on the direct grazing of forage resources with additional feeding, such as: grains, crop by-products, and stored forages such as hay or silage (Fulgueira et al., 2007). Cattle ranching is a major industry in Colombia that occupies about 38% of the land, employs 28% of the rural population and generates 3.5% of the country's GDP (Department of Energy & Climate Change, 2020). 82% of cattle farms in Colombia are owned by small-scale farmers, most of whom live in rural poverty and the prevalent method of grazing cattle on open pasture is environmentally harmful and economically unsustainable, providing many small farmers with low livelihoods. There is a need for sustainable and cheap precision agriculture techniques in Colombia which would help the rural farmers and lessen the burden on the climate and the pastureland.

This study explores the possible advantages of using Planet Scope data to predict forage biomass, dry matter digestibility, ash and crude protein concentrations. The study site is in Patía, Cauca department, Colombia which contains Brachiaria brizantha cv. Toledo, cv. Marandú, Brachiaria hybrid cv. Caymán, Mulato II and Megathyrsus maximus cv. Mombasa forage types. The aim of this project is to develop an approach for satellite-based monitoring and prediction of pasture quality and productivity, using the Planet Scope satellite datasets and machine learning algorithms. This study has two objectives:

- Analyse the image and field collected forage quantity relationships and develop a predictive model using the Random Forest algorithm.
- Demonstrate the use of remote sensing measurements and the derived predictive models to estimate forage characteristics across large areas.

2. Literature Review

2.1 Remote sensing techniques in Precision Agriculture

To use a familiar phrase, ‘if you cannot measure it you cannot manage it’, this is immensely relevant in the case of precision agriculture as the prime source is measurement data (Cox, 2002). Based on an evaluation of previous work, satellite driven grassland biomass estimation methodologies can be characterised into three sets: 1) using vegetation indices (VIs), 2) biophysical simulation models, and 3) machine learning algorithms (Ali et al., 2017). The use of machine learning has increased in the past decade as the results achieved were of superior quality than the traditional methods.

Remote sensing depends on the spatial and temporal resolution of the sensor used to relate between samples collected in the field and remotely sensed reflectance data like Sentinel-2 at 10-60m or Landsat Thematic Mapper, Enhanced Thematic Mapper+, Operational Land Imager at 30m spatial resolution (Raab et al., 2020; Zha et al., 2003b). Semi natural grasslands have a varied phenology that changes with seasons which require sensors with higher spatial resolution. TerraSAR-X (~3 m in strip map mode) has shown to be a significant data source to observe swath events for grassland areas (Schuster et al., 2011) and Worldview 2 data (saves multispectral data at 1.8 m spatial resolution) (Ramoelo, Cho, Mathieu, Madonsela, et al., 2015) has proven beneficial in estimating biochemical properties of grassland. The only constraint of these sensors is financial as they are operated by commercial companies which would be expensive in the long run. Landsat 8 Operational Land Imager (OLI) sensors offer free imagery with a spatial resolution of 30m (medium resolution) with 11 spectral bands and a temporal resolution once every 16 days (Landsat 8 « Landsat Science, 2020). Data from the Copernicus Sentinel-2 satellite missions is already used extensively for a range of agricultural applications. It offers high (10 m pixel size) to medium (20 and 60 m pixel size) spatial resolution data, combined with a higher spectral (13 bands) and temporal resolution (5 days) (Raab et al., 2020; Sentinel-2 - Missions - Sentinel Online, 2020). To estimate chlorophyll content the red-edge region in the electromagnetic spectrum is highly significant. The red edge variables are significantly correlated to nitrogen concentrations even at canopy level hence, the red-edge reflectance can therefore be related to the concentration of plant proteins (Adelabu et al., 2014; Clevers & Gitelson, 2012; Hoa, 2017; Mutanga & Skidmore, 2007; Tian et al., 2011). The use of red-edge bands improved the predictive ability to estimate biophysical parameters although Clevers et al., (2017) has shown that the Leaf Area Index (LAI) and chlorophyll concentration can be assessed using Sentinel-2 images without red-edge information. Planet Scope data does not have a red-edge band, but it is still feasible to get an accurate result. According to the study by Punalekar et al., (2018) the omission of red-edge bands, while creating 10 m LAI maps, did not affect the retrieval accuracy of pasture canopies, and the physically based approach exceeded the empiric NDVI approach.

Optical sensor-based approaches like Sentinel 2 are dependent on factors like illumination and cloud cover and rely on the sun's radiation hence, during rainy days there would be data gaps due to presence of clouds. Synthetic-aperture radar (SAR) sensors, provide many benefits for detecting change, provided that the longer wavelengths used by microwave sensors are not influenced by fog and haze (McNairn et al., 2014). TerraSAR-X and RADARSAT-2 was used to demonstrate that SAR can support the early estimates of production of corn and soybean, one of the two high value crops in Canada (McNairn et al., 2014). Breunig et al., (2020) delineated management zones in agricultural fields from south Brazil using Planet Scope data. This study confirmed the hypothesis that the cover-crop Above Ground Biomass (forage turnip, white oats, and rye) is linked with the cash-crop yield (soybean and maize) and can be assessed with the help of VIs. In a study by Ramoelo et al., (2012) using RapidEye imagery, vegetation indices gave poor results in estimating biomass, particularly during a time of peak productivity, due to known saturation problems. It gave the information that integrating vegetation index (SR54) and ancillary/environmental data is beneficial for model development in biomass variation estimation. Previous studies focused on the development of vegetation indices, but limited research on modelling algorithms existed, hence Random Forest regression algorithm was tested in this study.

According to the studies presented above, airborne imaging spectroscopy has been used to map the quality of the forage in grassland but carrying out manned airborne mission requires thorough preparation in advance. By comparison, an unmanned aerial vehicle (UAV) allows low-altitude images to be taken over wide areas with less effort and relatively low costs (Capolupo et al., 2015; Wijesingha et al., 2020). The development and miniaturisation of hyperspectral sensors helped in achieving more improvements in the study of grasslands (C. Zhang & Kovacs, 2012) which is why, the combination of UAVs and hyperspectral sensors has been applied in grassland mapping and monitoring and has reported its great capability to detect water stress and to estimate biophysical characteristics of grassland (Darvishzadeh et al., 2008; O. Mutanga et al., 2005; Schlerf et al., 2005).

2.2 Ensemble classifiers

The fundamental concept of ensemble learning is an understanding that every model will have limitations or weaknesses, and the goal of studying the ensemble is to control its strengths and weaknesses, contributing to the best possible overall decision-making (Brown, 2010). Ensemble methods constitute a large class of algorithms; this study focuses on using Random Forest regression model for the forage quality and quantity analysis. (Belgiu & Drăgu, 2016; Saini & Ghosh, 2017) explained the vital concept of ensemble classifiers, reviewing the different ensemble techniques (Boosting, Bagging, and Random Forest) particularly advantageous for remote sensing. According to Auret & Aldrich, (2012) two required and appropriate criteria must be met for an ensemble to be more accurate than its member: the members of the ensemble must have greater individual accuracies than random guessing, and the members of the ensemble must be diverse. Models are called diverse if they are uncorrelated to the errors they produce on unseen data. Machine Learning (ML) methodologies include a learning process with the goal of learning from training data to do a task. ML model success in a task is measured by a prediction model, which is enhanced over time with practice.

The study by (Yadav et al., 2019) demonstrated a comparative evaluation of the random forest regression (RFR), support vector regression (SVR) and artificial neural network regression (ANNR) algorithms used by Landsat-8 satellite data to assess LAI for crops. The high R^2 and low RMSE values between observed and predicted LAI indicated a high strength of the predictive regression algorithms. The maximum sensitivity of the normalised differential vegetation index (NDVI) with LAI was observed with RFR. SVM (Support Vector Machines), one of the prevalent machine learning algorithms, is widely used and has been explored in the remote sensing community for reasons of managing the problem of high dimensionality in a small range of training data and attaining high classification accuracy (Foody et al., 2006; Melgani & Bruzzone, 2004; Ustuner et al., 2016). Wei et al., (2017) demonstrated that the RF model yielded more reliable results than the ANN and SVM models when attempting to recover several growth stages of soybean Leaf Area Index. The brilliant performance of the RF model could be owing to its principle of 'majority vote,' which lessens the adverse effects of outliers. Moreover, structuring each decision tree on a subset of (mtry) bands is innately resilient to the overfitting issue (L. Wang et al., 2018). The RF method 's superiority to other methods has been consistent with recent studies (L. Wang et al., 2018; Wei et al., 2017; Yuan et al., 2017)

There are different approaches used to predict forage quality of grasslands: regression modelling is one of the most used non-parametric approaches/algorithms in research related to remote sensing. Numerous studies suggest that crop classification and mapping use the combination of machine learning methods and satellite imagery more frequently (Gislason et al., 2006; Ham et al., 2005; Rodriguez-Galiano et al., 2012). The RF algorithm is a non-parametric statistical procedure that can create regression or classification functions based on distinct or continuous datasets (Berhane et al., 2018). RF also has an ability to deal with complex relationships between predictors due to the noise and large amounts of data (Lu et al., 2019).

Many investigations reviewed by Belgiu & Drăgu (2016) have inspected the sensitivity of the RF classifier to sampling design and imbalanced training samples. It has been shown that the RF classifier is ideal for classifying hyperspectral data, where the dimensionality and strongly correlated data pose problems to other available classification procedures (Abdel-Rahman et al., 2013; Ismail & Mutanga, 2010; Onisimo Mutanga et al., 2012). It helped in getting a clear picture of the use of training samples in proportion to the area or getting an equal number of samples to each class. It also talked about the limitations when dealing with multi-modal input datasets (Breiman, 2001) because these classifiers assume a normal data distribution, which is rarely the case for remotely sensed data. According to Raab et al., (2020) random forest regression algorithm was used to relate field and the satellite data at a study site in Germany. The findings suggested that Sentinel 2 data was more efficient than Sentinel 1, and Sentinel 1 and 2 combined respectively, in accurately predicting the forage quality and quantity. The study also talked about the possible things to keep in mind for future studies e.g., in order to reduce the geolocation error effect, careful selection of the plot size is important. Overall, the optimized subset of the predictor dataset increased the random forest regression model performance and the value of the red-edge region (Abdel-Rahman et al., 2013; Hunt et al., 2019; Ramoelo et al., 2015) to predict biophysical parameters was confirmed.

Using RF in combination with recursive feature elimination, Meyer et al., (2017) tested whether complete hyperspectral data is required, or whether multispectral (QuickBird, RapidEye and WorldView-2 in this case) information is enough to correctly estimate proxies for pasture degradation in Qinghai-Tibet Plateau (QTP). It also emphasized the issue of scale differences between the locally taken samples and the spatial resolution of the images and found that hyperspectral data had no benefit over multispectral vegetation cover and AGB modelling data. The sensitivity of RF to highly correlated variables was also verified by Zhen-wang et al., (2017) to examine its stability. The results exhibited that mtry, a vital RF parameter, has a minor consequence on the performance of RF. Two approaches were used to reduce the predicted variables; one was based on the Variable Importance Value, and the other was based on the Principal component analysis (PCA). It was recommended to use the Variable Importance Value method rather than the PCA method; even if the variable reduction hardly improved the RF performance, it showed better potential for predictions with multiple inputs and enormous data.

Shimizu et al., (2020) also confirmed the previous studies where the ability of Sentinel-2 short-wave infrared and red-edge bands for structural forest mapping had been proved (Astola et al., 2019; Mauya et al., 2019). One of the reasons why Sentinel-2 is more accurate than PlanetScope, which only has four spectral bands was confirmed by the high variable importance of variables linked to short-wave infrared in Sentinel-2. Although the Planet Scope data 's original spatial resolution is higher than that of Sentinel-2, the RF models were more accurate with Sentinel-2. Seeing the minor differences between the prediction accuracies of Sentinel-2 and PlanetScope, use of Planet Scope data was encouraged for near-real-time monitoring of forest traits. In the study by Lu et al., (2019), the PLS and RFR models were implemented as models of regression for estimating chlorophyll content from multispectral and hyperspectral imagery. Using RF regression produced better results because many regression trees are constructed independently in the model using training data sub-sets, and thus it is not prone to data noise. Moreover, all predictor variables are considered for growth of trees at each node and thus model output is optimised in the case of RF regression. Like other studies (Koppe & li, 2010; Meyer et al., 2017; Tong & He, 2017b) it was also found that the hyperspectral image with 325 bands did not perform better than the red-edge image with five bands because its spectral bands are mostly correlated, it does not have any more predictive capacity to estimate chlorophyll content.

This study is one of its kind in demonstrating the potential of Planet Scope data used to monitor forage characteristics to enhance livestock management at low costs. The past studies have used Planet Scope satellite data in a combination with Unmanned Aerial System (sUAS) (Liu et al., 2019) or comparison with different satellites like Landsat, MODIS and Sentinel (Gašparović et al., 2018; Houborg & McCabe, 2018; Shimizu et al., 2020); studies just using Planet Scope dataset were about determining aboveground biomass (AGB) (Breunig et al., 2020; dos Reis et al., 2020; Miller et al., 2019), estimating vegetation phenology (Cheng, 2019), comparison of various spectral indices (Hoa, 2017) or evaluating the relationship between stem water potential and vegetation indices (Helman et al., 2018).

3. Methodology

3.1 Study area

Cauca is a Department of Colombia, located in the southwestern part of the country, facing the Pacific Ocean to the west, the Valle del Cauca Department to the north, Tolima Department to the northeast, Huila Department to the east, and Nariño Department to the south (Departamento de Valle Del Cauca (Colombia) - EcuRed, 2020) (Figure 3.1).

The study area, Patía is a Colombian municipality located in the southern province of Cauca department with an administrative centre known as El Bordo. Patía is located on the Panamerican road, about 82 km south of Popayán, capital of Cauca department. It has an elevation of 990m and a population of 37,781. The territory of the municipality is crossed by the river Patía. The municipality borders to the east by La Sierra and Bolvar, to the west by Argelia and Balboa, to the north by El Tambo and La Sierra, and to the south by Sucre and Mercaderes. Its geographical coordinates are 2.117° N and 76.983° W. The average temperature is 23°C with an average annual precipitation of 2,171 mm.

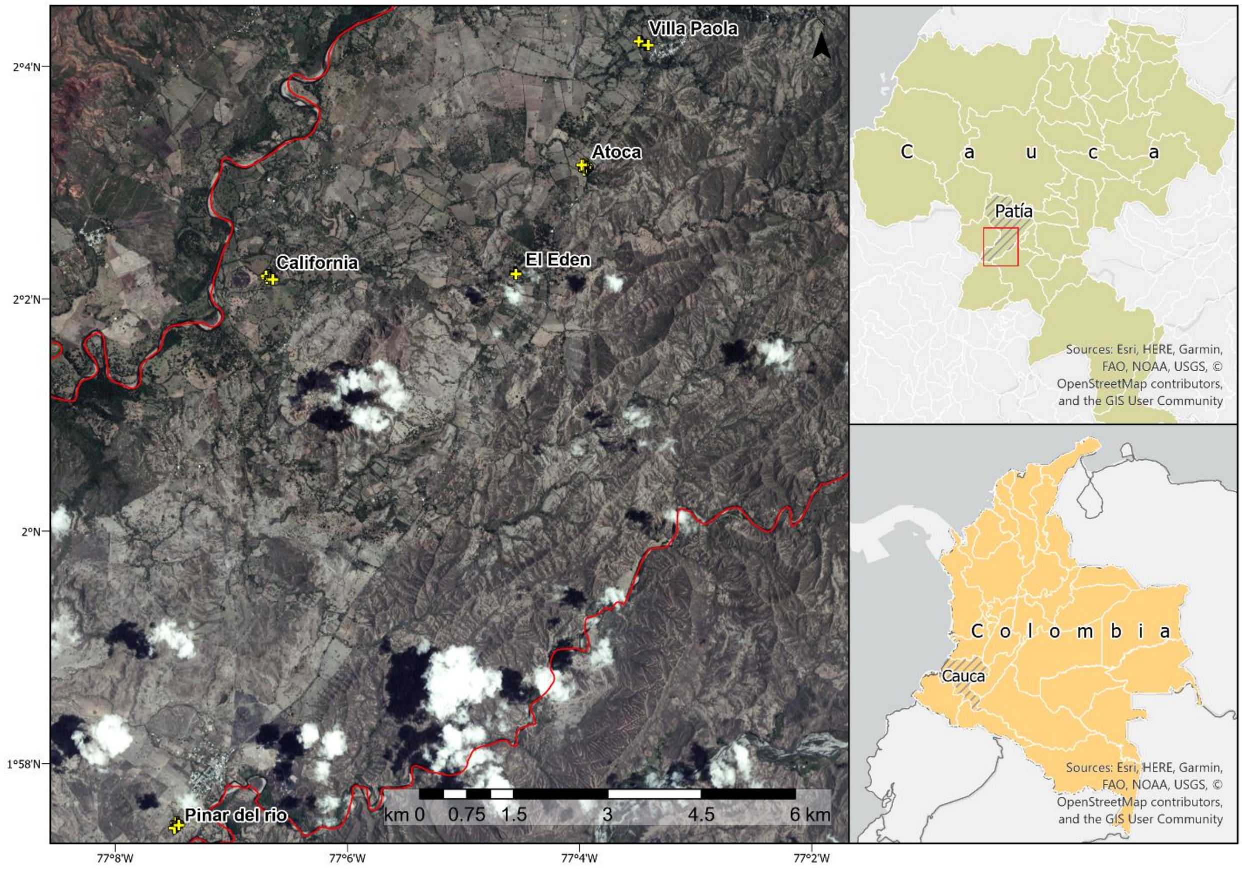


Figure 3.1. Location of the study site Patía. The location of the study site in Cauca, Colombia is highlighted with hatched shading (top right corner). The five sampling locations are marked with yellow crosses and labelled as per their locations. The base imagery is a mosaic of the three passes dated 18-09-2018 of the Planet Scope satellite.

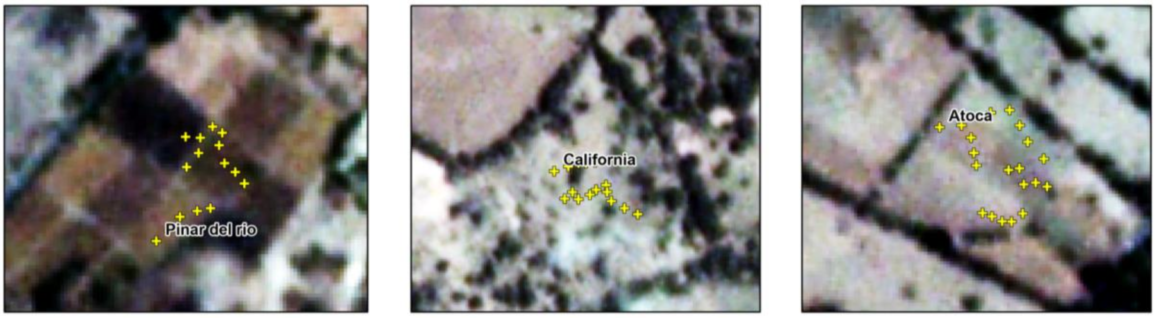


Figure 3.2. Three sampling locations out of the five marked in yellow namely Pinar del Rio, California and Atoca (left to right)

3.2 Field Data

Ground truth data collection occurred across three dates (17 – 19th Sept 2018) in the study area and was carried out by a team of researchers from the University of Glasgow and the International Centre for Tropical Agriculture (CIAT). Forage samples were collected using a 0.5 x 0.5 m quadrat and analysed for fresh and dry weight biomass, crude protein, ash percentage, and in-vitro dry matter digestibility. Prior to removal, the average forage height was measured using a ruler and the NDVI at 1 m height above the quadrat using a GreenSeeker™ handheld optical sensor unit. The location of each quadrat was recorded using a South G1 differential GPS receiver. The sampling locations and number of samples are presented in table 3.1; and figure 3.4 gives a summary of the forage quantity and quality parameters provided by the field dataset.

Figure 3.3. Photographs from one of the sampling sites called Atoca in Figure 3.1 and Figure 3.2. Photographs show the species (A) Angleton, (B) Mulato II and (C) Toledo.

Table 3.1. Sampling dates of the field dataset and corresponding satellite data acquisitions used in this study

Field Sampling Date	Location	Number of samples	PlanetScope Acquisition Date
18-09-2018	El Eden	5	
18-09-2018	California	15	
18-09-2018	Villa Paola	10	18-09-2018
19-09-2018	Atoca	20	
19-09-2018	Pinar del Rio	15	

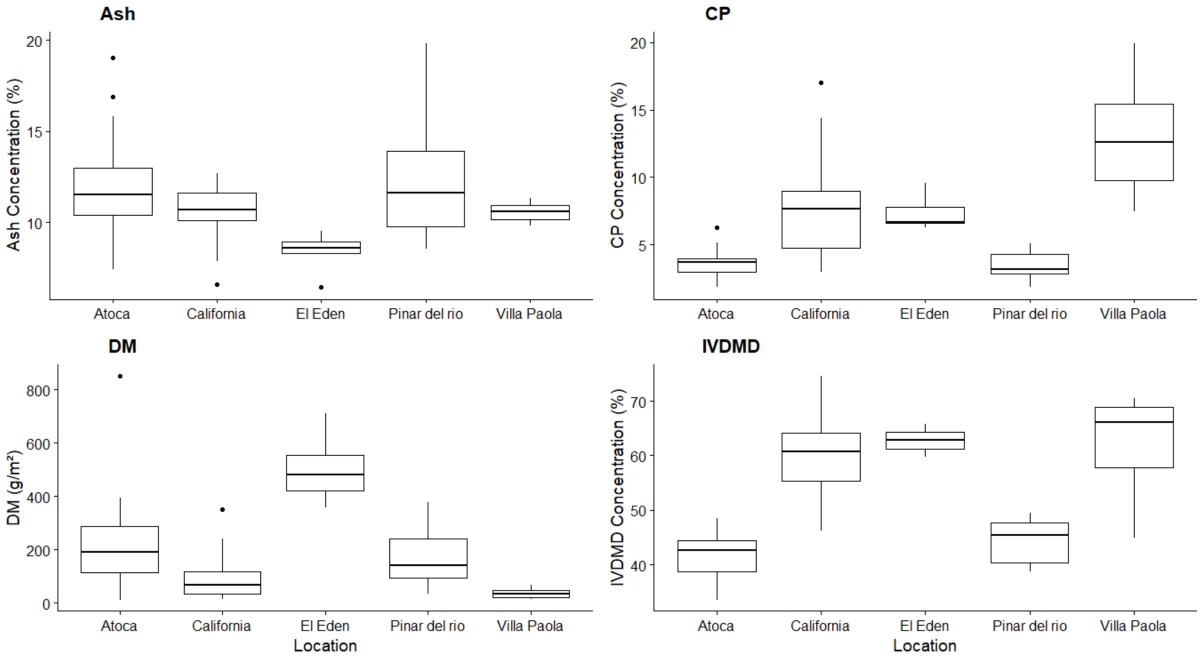


Figure 3.4. Grassland forage quantity and quality data used in this study. The respective sampling dates are shown in Table 3.1.

3.3 Satellite data and pre-processing

Planet operates with more than 175 Planet Scope satellites and collects multispectral (MS) imagery in 4 bands with a spatial resolution of ~3 m and a collection capacity of 300 million square km per day. Planet, an aerospace corporation, develops and manages the largest constellation of small imaging satellites, Planet Scope (PS), also known as Cubesat or Dove (Asner et al., 2017). The data was received as level 3B product (Planet Scope Analytic Ortho Scene product is orthorectified, 4-band B-G-R-NIR Imagery with geometric, radiometric and atmospheric correction) suitable for analytic and visual application. The scenes were provided already orthorectified to <10 m RMSE positional accuracy and projected to UTM/WGS84 cartographic projection (Mudereri et al., 2019). Three scenes from the same date (18/09/2018) were mosaicked to cover the whole study area.

Table 3.2. Spectral and spatial specifications of the Planet Scope constellation

Band	Band Name	Central Wavelength (nm)	Spatial Resolution (m)	Bandwidth (nm)
1	Blue	485	3	60
2	Green	545	3	90
3	Red	630	3	80
4	Near Infrared	820	3	80

3.4 Vegetation Indices

Information on the quality of food used to be collected by biomass sampling for decades; biomass sampling helped inform the vegetation communities about the effects of herbivores (Borowik et al., 2013). Most VIs are easy to measure and can minimise variability due to site-specific settings, such as bare soil, angle of illumination, or atmosphere (Raab et al., 2020). In total, 30 predictor variables were available for the grassland forage quantity and quality random forest regression models, including 4 multispectral bands, 16 vegetation indices and 10 Simple Ratios as seen in Table 3.3. The indices were computed, using the freely available R statistical programming environment and ArcGIS Pro 10.5 was used to cross check the values.

Table 3.3. Summary of vegetation index expression (*Broadband Greenness*, 2020)

Vegetation Index	**Equation**	**References**
Normalized Difference Vegetation Index (NDVI)	$\mathrm{NDVI} = \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red})}$	(Rouse et al., 1974)
Soil Adjusted Vegetation Index (SAVI)	$\mathrm{SAVI} = \frac{1.5 * (\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + \mathrm{Red} + 0.5)}$	(A. R. Huete, 1988)
Modified Soil Adjusted Vegetation Index 2 (MSAVI2)	$\mathrm{MSAVI2} = \frac{2 * \mathrm{NIR} + 1 - \sqrt{(2 * \mathrm{NIR} + 1)^2 - 8(\mathrm{NIR} - \mathrm{Red})}}{2}$	(Qi et al., 1994)
Green Optimized Soil Adjusted Vegetation Index (GOSAVI)	$\mathrm{GOSAVI} = \frac{\mathrm{NIR} - \mathrm{Green}}{\mathrm{NIR} + \mathrm{Green} + 0.16}$	(Ravi Prakash Sripada, 2005)
Ratio Vegetation Index (RVI)	$\mathrm{RVI} = \frac{\mathrm{Red}}{\mathrm{NIR}}$	(Xue & Su, 2017)
Infrared Percentage Vegetation Index (IPVI)	$\mathrm{IPVI} = \frac{\mathrm{NIR}}{\mathrm{NIR} + \mathrm{Red}}$	(Crippen, 1990)
Difference Vegetation Index (DVI)	$\mathrm{DVI} = \mathrm{NIR} - \mathrm{Red}$	(Tucker, 1979)

Simple Ratio (SR)	$\mathrm{SR} = \frac{\mathrm{NIR}}{\mathrm{Red}}$	(Birth & McVey, 1968)
Green Ratio Vegetation Index (GRVI)	$\mathrm{GRVI} = \frac{\mathrm{NIR}}{\mathrm{Green}}$	(Ravi P. Sripada et al., 2006a)
Enhanced Vegetation Index (EVI)	$\mathrm{EVI} = 2.5 * \frac{(\mathrm{NIR} - \mathrm{Red})}{(\mathrm{NIR} + 6 * \mathrm{Red} - 7.5 * \mathrm{Blue} + 1)}$	(A. Huete et al., 2002)
Leaf Area Index (LAI)	$\mathrm{LAI} = 3.618 * \mathrm{EVI} - 0.118$	(Boegh et al., 2002)
Green Ratio Vegetation Index (GRVI)	$\mathrm{GRVI} = \frac{\mathrm{NIR}}{\mathrm{Green}}$	(Ravi P. Sripada et al., 2006b)
Green Leaf Index (GLI)	$\mathrm{GLI} = \frac{(\mathrm{Green} - \mathrm{Red}) + (\mathrm{Green} - \mathrm{Blue})}{(2 * \mathrm{Green}) + \mathrm{Red} + \mathrm{Blue}}$	(Louhaichi et al., 2001)
Modified Simple Ratio (MSR)	$\mathrm{MSR} = \frac{\left(\frac{\mathrm{NIR}}{\mathrm{Red}}\right) - 1}{\left(\sqrt{\frac{\mathrm{NIR}}{\mathrm{Red}}}\right) + 1}$	(Chen, 1996)
Non-Linear Index (NLI)	$\mathrm{NLI} = \frac{\mathrm{NIR}^2 - \mathrm{Red}}{\mathrm{NIR}^2 + \mathrm{Red}}$	(Goel & Qin, 1994)
Visible Atmospherically Resistant Index (VARI)	$\mathrm{VARI} = \frac{\mathrm{Green} - \mathrm{Red}}{\mathrm{Green} + \mathrm{Red} - \mathrm{Blue}}$	(Gitelson1 et al., 2002)
Renormalized Difference Vegetation Index (RDVI)	$\mathrm{RDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\sqrt{(\mathrm{NIR} + \mathrm{Red})}}$	(Roujean & Breon, 1995)
Green Normalized Difference Vegetation Index (GNDVI)	$\mathrm{GNDVI} = \frac{(\mathrm{NIR} - \mathrm{Green})}{(\mathrm{NIR} + \mathrm{Green})}$	(Gitelson & Merzlyak, 1998)

3.5 Statistical analysis

3.5.1 Prediction Assessment

The random forest regression algorithm was used to evaluate the relationship between grassland forage quantity and quality and remote sensing resultant datasets. The root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) between measured and predicted values were used to assess the model performance. RMSE and MAE were calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{n}(\mathrm{Yi} - \hat{\mathrm{Y}}\mathrm{i})^2}{n}} \qquad (1)$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|\mathrm{Y_i} - \widehat{\mathrm{Y_i}}|}{n} \qquad (2)$$

where Yi is the measured value, Ŷi the predicted value of cases i and n equals the number of observations. Relevant variables were first defined using the random forest variable importance feature. In random forest model, a variable is considered important if omitting it from the predictor variable list increases the OOB error (Liaw & Wiener, 2002). The most widely used variable importance metric for random forest regression is permutation based MSE (Mean Square Error) reduction (Strobl et al., 2008), called permutation importance. Permutation importance is constructed as follows from the random forest MSE: Out-bag observations, i.e. observations not used to build the tree, can be used in each tree to determine the MSE from that tree. The overall development of the forest can be obtained from the individual trees by an adequate combination of the MSE estimates. A variable's contribution is calculated by randomly permutating the observations for that variable and measuring the difference between the prediction performance with the real and the permuted variable values (Grömping, 2015; Prasad et al., 2006).

3.5.2 Random Forests

Unlike bagging trees, RF grows its trees with a randomly selected subset of the number of predictors at each splitting node (mtry), and the tree can grow fully without pruning. Each tree in the RF is independently grown to its maximum size based on the repeated cross-validation (ten-fold, 100 repeats) procedure from the training dataset (approximately two-thirds), and the remaining one-third of the samples are randomly left out; these are called the out of-bag (OOB) samples, which are used to compute an unbiased OOB error rate and variable importance (measured by calculating the percent increase in the mean square error when the OOB data for each variable are permuted) (Breiman, 2001; Prasad et al., 2006). At each binary split, the predictor that produces the best split is chosen from a random subset (mtry) of the complete predictor set (p), and mtry is recognised as the key tuning parameter of RF and should therefore be optimized (Zhen-wang et al., 2017). Using the out-of-bag samples, the prediction error (OOB error) for each individual tree is obtained using the following equation:

$$\mathrm{ERROR}_{\mathrm{OOB}} = \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Y_i} - \widehat{\mathrm{Y_i}}\right)^2 \qquad (3)$$

where, $\hat{Y}i$ is the predicted output of an OOB sample, Yi is the actual output and n is the total number of OOB samples. All machine learning approaches have configuration parameters called tune parameters or hyperparameters to enhance predictive modelling algorithms efficiency (Raschka, 2018). For random forest regression model, two tune parameters need to be determined: mtry, which is the number of randomly selected variables and ntree, which represents the number of trees to grow. For this study the number of trees parameter was set to vary from 500 to 1000 in the RF model and the mtry parameter value was tuned using the repeated cross-validation (ten-fold, 100 repeats) technique. The mtry parameter value was set between 1 and 8 and the optimum mtry parameter for each model was recognised with the help of the smallest RMSE value.

Finally, the spatial distribution of IVDMD, CP, Ash and DM was predicted using the best variable combination. All final maps were averaged over 30 predictions (VI+ SR+ Bands). All analyses were carried out within the R statistical programming environment (The R Core Team, 2020) using the packages ranger for RF regression (Wright & Ziegler, 2017), caret (Kuhn, 2020) for cross-validation and permutation, and raster (Hijmans, 2019) for the spatial predictions.

4. Results

4.1 Selection of predictor dataset and validation

RF models were built using different combinations of the predictor variables, including the seven types of predictors, the bands and vegetation indices (Bands+ VI), bands and Simple Ratios (Bands+ SR), the vegetation indices and simple ratios (VI+ SR), the bands, vegetation indices and simple ratios (Bands+ VI+ SR), the bands, the simple ratios (SR) and the vegetation indices (VI) individually. The lowest RMSE at the optimum mtry values for the combination (Bands+ VI+ SR) dataset is illustrated in Figure 4.1. The high R^2 values indicated a good match between observed and predicted IVDMD concentrations. This was assisted by relatively low RMSE values in comparison to the range of IVDMD concentrations measured in the field (Table 4.1). The small R^2 and $RMSE_{sd}$ values after 100 repetitions further maintained good model performances. For Ash and DM, lower R^2 values were achieved compared to IVDMD and CP. The estimated RMSE values were moderately higher for DM, bearing in mind the respective observed range of the data. For both CP and DM, the optimum number of randomly selected variables (mtry) parameter remained one, which was selected based on smallest RMSE value.

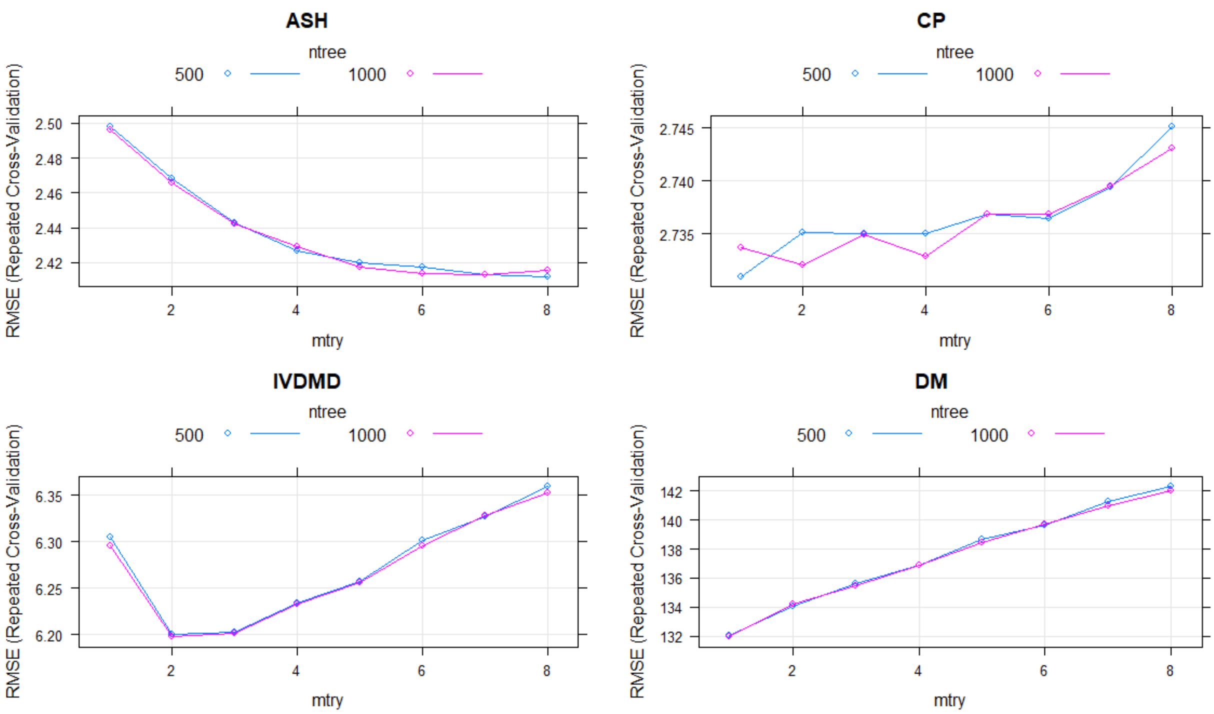


Figure 4.1. The optimum mtry parameter for each model identified with the help of the smallest RMSE value from the Planet Scope predictor dataset (VI+ SR+ Bands).

Table 4.1. Performance of the predictor dataset (VI+ SR+ Bands) estimated using random forest regression. Predictors were iteratively removed based on variable importance. The values are means of 100 repetitions of a 10-fold cross-validation (Abbreviations: CP, crude protein concentration; IVDMD, in vitro dry matter digestibility; DM, standing biomass dry matter weight).

	ASH (%)	**CP (%)**	**DM (g/m²)**	**IVDMD (%)**
Max observed	19.80	19.89	851.14	74.61
Min observed	6.44	1.86	12.61	33.54
mtry	8	1	1	2
ntree	500	500	1000	1000
RMSE	2.41	2.73	131.74	6.2
$RMSE_{sd}$	0.95	1.17	73.59	1.78
MAE	1.88	2.13	104.80	5.34
MAE_{sd}	0.75	0.85	44.62	1.51
R^2	0.38	0.69	0.49	0.74
R^2_{sd}	0.30	0.27	0.33	0.19
% Var explained	10.99	49.69	23.06	64.82

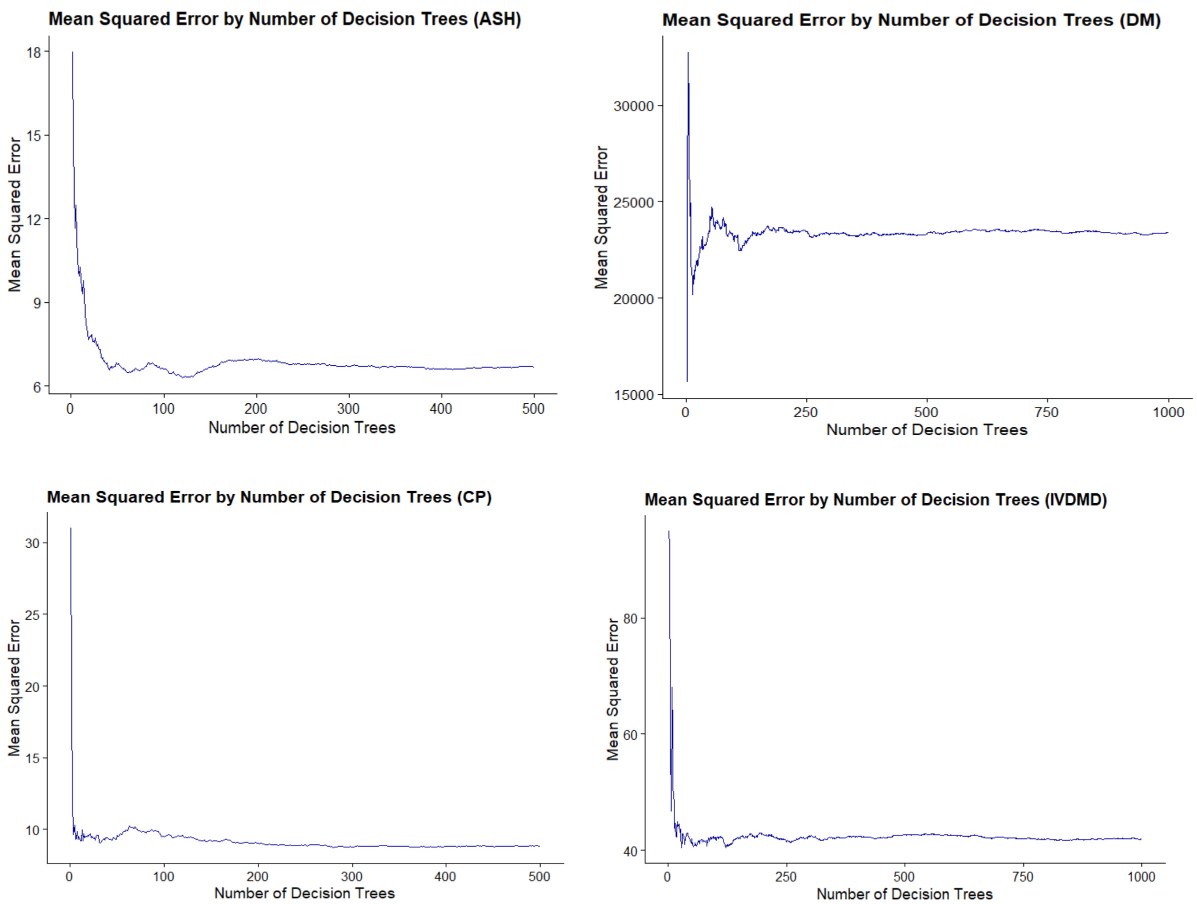


Figure 4.2. RF parameter optimization of the out-of-bag (OOB) error variation altering with the number of trees (ntree).

Before implementing the algorithm using the predictors, two important user-defined parameters of RF, ntree and mtry, were optimized to minimize the generalization error. Figure 4.2 shows the OOB error in response to the number of trees from 1 to 1000 using the tuned mtry (1 to 8) set during customisation of RF. When the trees grew from 1 to 100, the OOB error decreased rapidly and reached a minimal value at the point between 100 and 200 trees, but a fluctuation and increase in the OOB error was observed until approximately 500 trees. The OOB error remained consistent and did not indicate obviously better performance after that, so ntree=500 was chosen for Ash and CP in the study. The ntree values for DM became stable around 250 trees with minor fluctuations and the same type of result was seen for IVDMD where optimum ntree was 1000.

4.2 Variable importance

Permutation-based variable importance for the combination (VI+ SR+ Bands) was used to find important predictor variables for Ash, CP, IVDMD and DM (Figure 4.2). The key parameter mtry for RF was determined by a random search with a repeated (100 times) 10-fold cross validation procedure on the calibration dataset. For IVDMD and CP, the most important variable was NLI. The R^2 values for IVDMD were the lowest for the combination of VI+ SR and just SR; SR combination also had the highest RMSE. The highest R^2 and lowest RMSE was recorded when VI+ Bands combination was used for IVDMD and the most important variable was DVI. The index MSR contributed to the model performance of DM when VI+ SR+ Bands was used. The lowest RMSE (130.28) and highest R2 (0.50) was found for the Bands+ VI combination; the most important variable was SAVI. The highest RMSE (139.51) was calculated for the SR combination and the lowest R^2 (0.43) was found for Bands. VI+ SR+ Bands and VI+ SR combinations had MSR as the most important variable.

For Ash, the Red band was found to be the most appropriate variable for VI+ SR+ Bands, Bands+ VI and Bands. The lowest RMSE (2.41) was observed for Bands+ VI and the highest R^2 (0.37) was found for multiple combinations as seen in Table 4.2. Simple ratios and red band seem to be dominating as most important variable for all Ash combinations. Vegetation indices and the NIR band was the most important variable for CP. For IVDMD and DM, vegetation indices were more vital than for CP and Ash. The general contribution of Planet Scope blue and green bands was low compared to simple ratios or vegetation indices. The red and NIR bands were among the top ten important predictor variables in Ash, CP and IVDMD but no band was among the top 10 in the case of DM. The highest R^2 value among all the combinations was for Bands+ VI for IVDMD. Overall, the combination of Bands+ VI produced the best results in terms of RMSE and R^2 values. In the study by Guo et al., (2010), region of visible wavelength was considered important and the red region (600–700 nm) was associated with protein and ash. They also noticed no significant link between chemical content and spectral characteristics beyond the near infrared energy spectrum. This finding approved with (Tucker, 1977) suggesting that the most important for biophysical characterisation of grassland is the visible wavelength region.

Table 4.2. RF models built using different combinations of the predictor variables.

		Ash	CP	IVDMD	DM
VI+ SR+ Bands	RMSE	2.41	2.73	6.19	131.74
	R^2	0.34	0.69	0.73	0.49
	Most Important Variable	Red	NLI	NLI	MSR
Bands+ SR	RMSE	2.42	2.74	6.07	133.79
	R^2	0.37	0.70	0.74	0.47
	Most Important Variable	G / R	NIR	NIR	B/ NIR
Bands+ VI	RMSE	2.41	2.73	5.99	130.28
	R^2	0.37	0.69	0.75	0.50
	Most Important Variable	Red	RDVI	DVI	SAVI
VI+ SR	RMSE	2.44	2.77	6.62	135.25
	R^2	0.37	0.68	0.69	0.47
	Most Important Variable	R/ G	GNDVI	NLI	MSR
Bands	RMSE	2.53	2.83	6.03	135.61
	R^2	0.34	0.68	0.74	0.43
	Most Important Variable	Red	NIR	NIR	Blue
VI	RMSE	2.44	2.77	6.39	136.37
	R^2	0.37	0.68	0.72	0.46
	Most Important Variable	VARI	RDVI	DVI	EVI
SR	RMSE	2.45	2.85	7.34	139.51
	R^2	0.36	0.66	0.62	0.44
	Most Important Variable	R/ G	B/ NIR	NIR/ B	B/ NIR

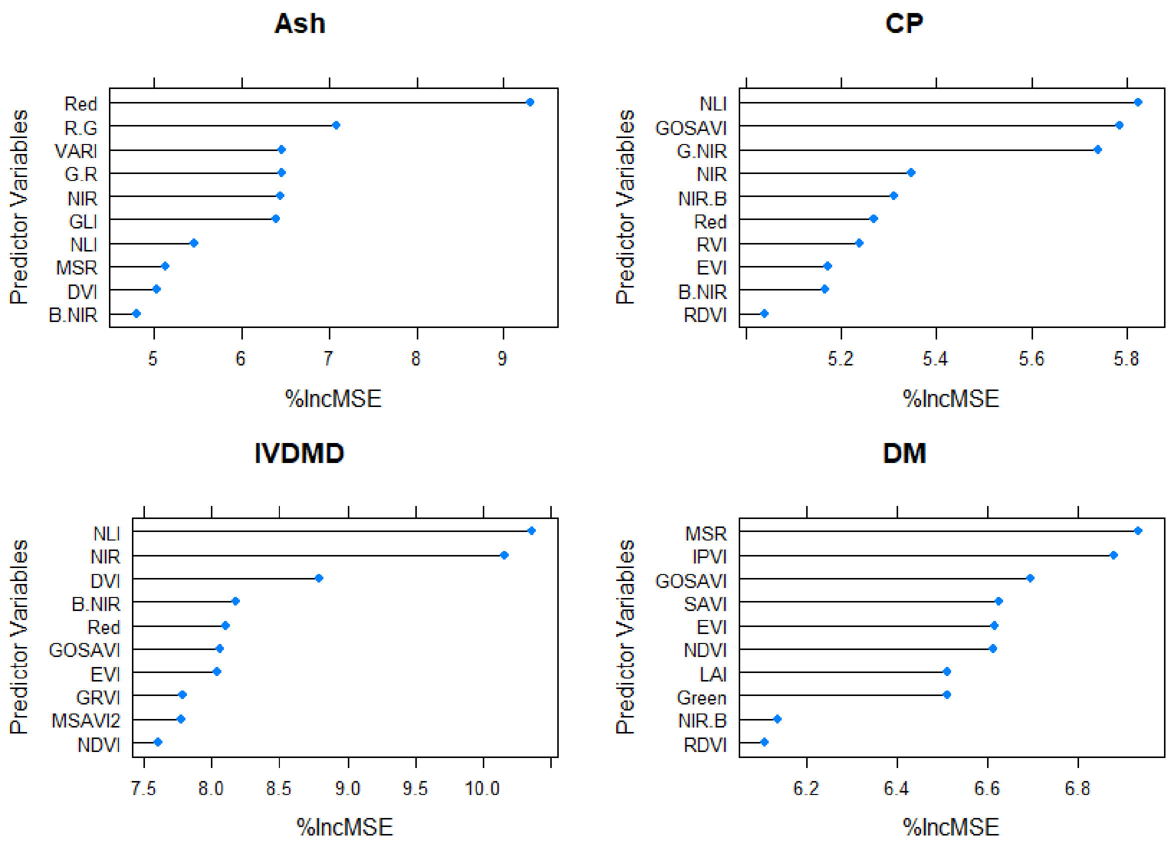
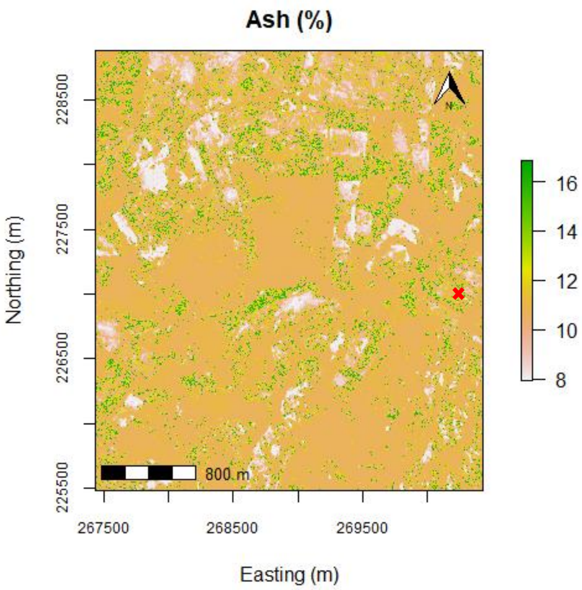


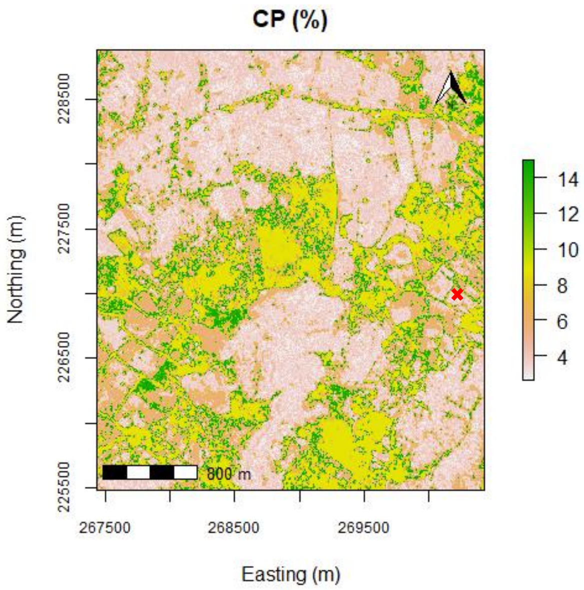
Figure 4.3. Variable importance for VI+ SR+ Bands calculated in terms of Mean Decrease Accuracy (%IncMSE) created by permuting the values of each variable of the test set, recording the prediction and comparing it with the unpermuted test set prediction of the variable (Top 10 variables shown).

4.3 Spatial prediction

Figure 4.4 shows the spatial predictions for Ash, CP, IVDMD, and DM derived from the optimised predictor variables (VI+ SR+ Bands) by implementing the respective RF model. This prediction was done using Red band for Ash, NLI for CP and IVDMD, and MSR for DM (check table 3.3 for abbreviations of VIs) due to low RMSE values (Table 4.2). For illustration purposes, the figures and subsequent descriptions are provided for the area neighbouring the sample location Atoca (marked in red). The results for September 2018 are presented. Areas with Ash concentration less than 8 % were clearly evident on the map and there is little distinction between areas as most values lie between 10% - 12%; high Ash concentrations show that the forage might be contaminated (White, 2018). The area surrounding Atoca has a mixed concentration of Ash (~10 %-14%. The mean predicted Ash concentration was about 11.34 % which matches with the values of the plot shown in Figure 3.4. As in the case of CP, the mean concentration was 6.54 %. A distinct differentiation in areas was seen with CP concentration. Similar to the predictions of CP, IVDMD showed a spatial variation, where the map seems divided into values above and below 55 %. The mean predicted IVDMD concentration was 50. 8%. A similar distinction was made for the DM predictions, which had a mean value of 187.28 g/ m^2. For DM, the areas surrounding Atoca mostly show values below 300 g/ m^2.

In summary, more pronounced differentiation was observed for CP, IVDMD and DM whereas Ash showed a relatively even distribution. Areas with low Ash concentration and low DM values were related to high CP and IVDMD concentration and vice versa.





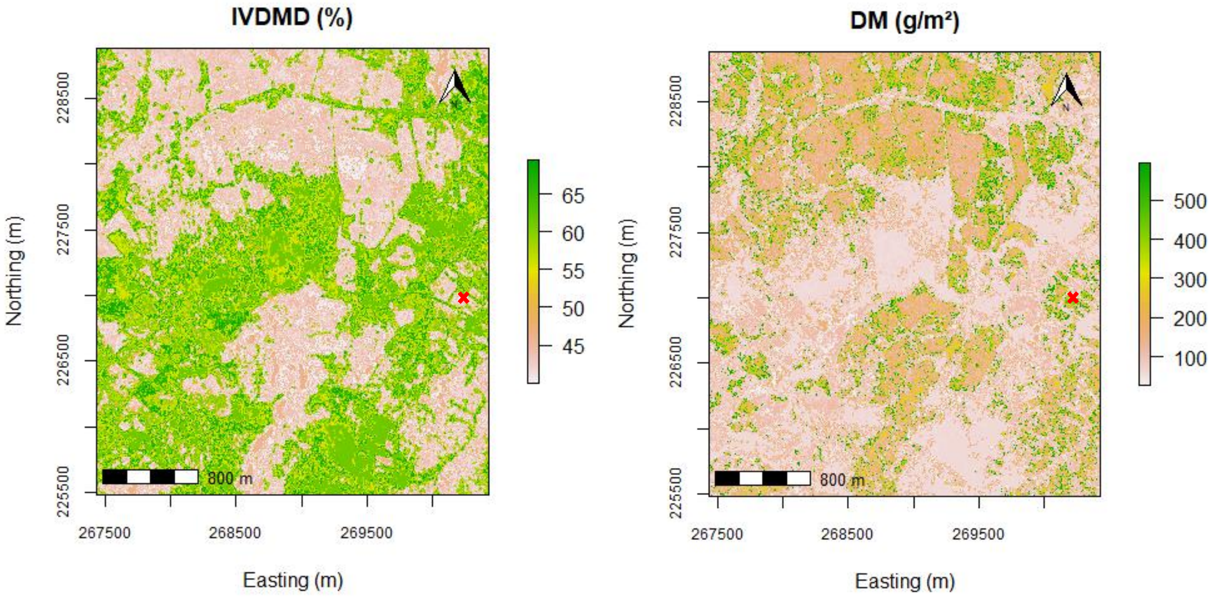


Figure 4.4 . Spatial predictions of Ash, CP, IVDMD and DM for September 2018 using random forest regression, averaged over 100 repetitions. The illustrations are shown for the area neighbouring the sampling location Atoca (marked in red) of Figure 3.1.

5. Discussion

This study used the RF model to assess the value of different predictors and estimate the forage quality and quantity parameters, and it attained good results both in terms of model accuracy and spatial patterns. It also shows that forage quality indicators can be mapped with high accuracy for some parameters, as R^2 values of the regression model (VI+ SR+ Bands) was 0.74 (sd = 0.19) for IVDMD and 0.69 (sd = 0.27) for CP but 0.38 (sd = 0.30) for Ash. For the grassland forage quantity indicators (DM), lower R^2 values were obtained R^2 = 0.49, sd = 0.33.

5.1 Planet Scope data for grassland forage quantity and quality prediction

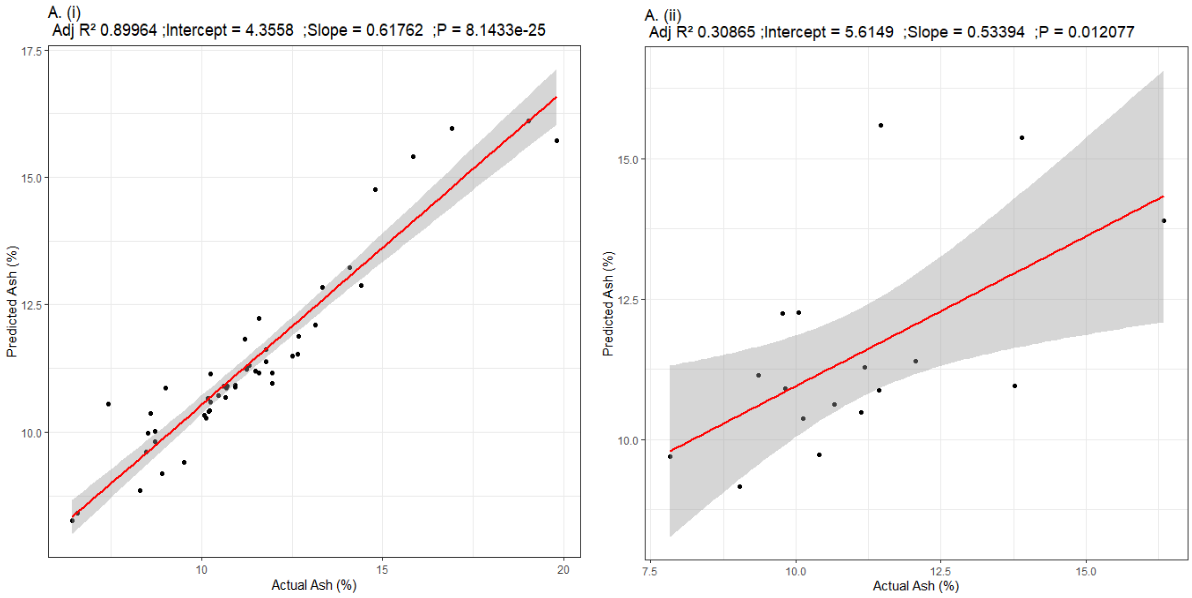
Planet's nano-satellite constellation of CubeSats provides an unparalleled ability to track vegetation dynamics more regularly than ever before with enhanced spatial accuracy (Helman et al., 2018; Miller et al., 2019). The temporal and spatial resolution of commonly used remotely sensed optical data , in particular medium and large-scale remotely sensed imagery, has been a barrier to beneficial fine-scale pasture monitoring in intensively managed fields (Otgonbayar et al., 2019; J. Wang et al., 2019; B. Zhang et al., 2015). The images usually have a moderate cloud coverage during this time (mid-September), and it might be concluded that the high temporal resolution of the Planet Scope data was critical to achieve maximum coverage and thus a cloud-free mosaic. Even though the quantity of research concerned with the assessment of biophysical parameters using only Planet Scope data for forage is limited, the methodology to estimate the random forest regression was similar.

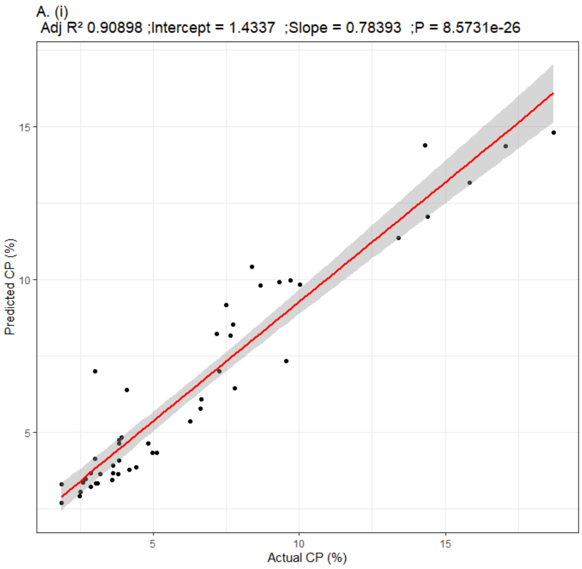
5.2 Optimization of the predictor dataset and important variables

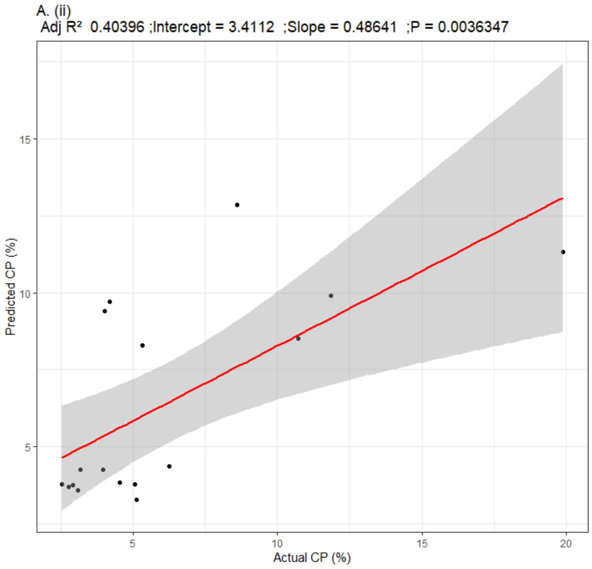
As stated in the study by (Breiman, 2001; Grömping, 2015; Prasad et al., 2006) RF has two procedures of variable importance. The first is based on a mean squared error (MSE) and relates to the predictive accuracy of the data's out-of-bag component after each predictor variable has been permutated. The difference between the two MSEs is then averaged across all trees and normalised by standard error. The second calculation is the same as for bagging trees, which is measured based on the data used to grow trees. Consequently, the assumption is based on overfitted models. One to one relationship between actual and predicted biomass using RF regression models is shown in Figure 5.1. For each model, R^2, Intercept, Slope and P value were reported. The use of random forest and SR provided poor prediction for biomass compared to VI+ Bands which show R^2 values 0.88 and 0.57 for training and testing data respectively. For the VI+ Bands combination, the R^2 values for CP, IVDMD and Ash using the testing data were 0.40, 0.42 and 0.31 respectively. It can be easily noted that the R^2 testing values of Ash were the lowest among DM, IVDMD and CP. The results are promising as they show high R^2 values for the training dataset; the values are more scattered when using the testing values because of the small amount of sample numbers used as compared to the training dataset.

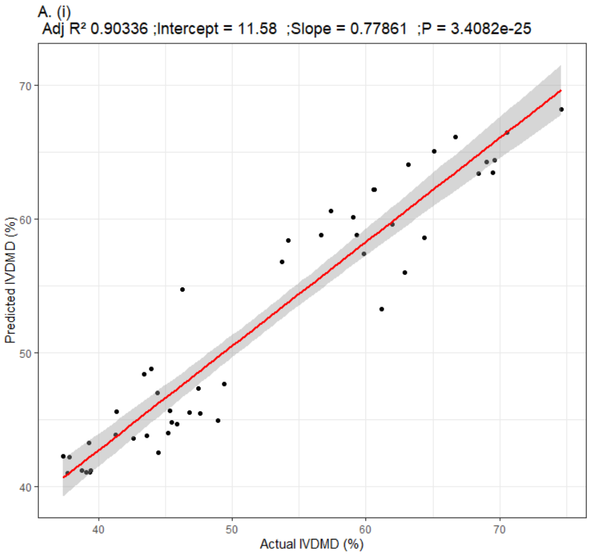
Studies by (Duro et al., 2012; Immitzer et al., 2012; Liaw & Wiener, 2002; H. K. Zhang & Roy, 2017) have shown that adequate efficiency is achieved with the default parameters. Liaw & Wiener (2002) stated in their paper that a large number of trees can provide a consistent outcome of variable importance. It was brought into light by Breiman (2001) that having more than required number of trees does not affect the model although it may be pointless. The tuning parameters for this study, to find the optimal RF regression model, a range of values for both parameters were tested and evaluated: ntree = 500 and 1000; mtry = 1:8. The process of variable selection in the case of random forest machine learning must be seen as a significant and necessary processing step (Gregorutti et al., 2017).

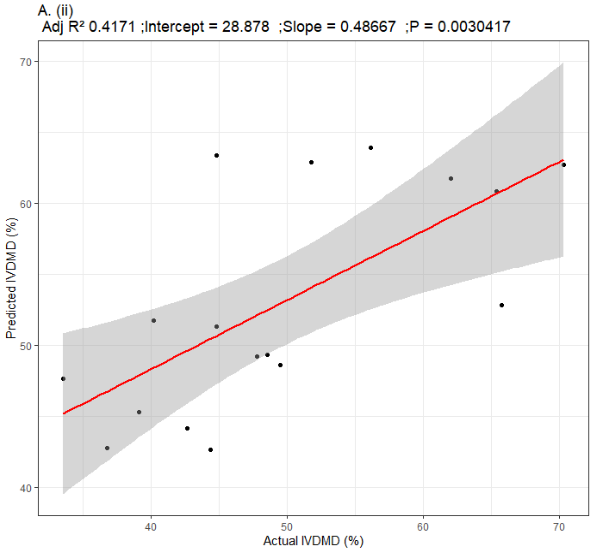
As the chlorophyll concentration is related to the crude protein concentration (Gáborčík, 2003; Rincón Castillo et al., 2019), the high value of VI+ Bands combination was due to the fact that near-infrared band and the red band was present in each vegetation indices used. For real world applications, the choice of a new VI must be made with caution by comprehensively evaluating and examining the advantages and disadvantages of existing VIs, and then integrating them to be implemented in a particular setting. This is how the use of VIs can be adapted to particular applications, instruments used and various platforms (Xue & Su, 2017).

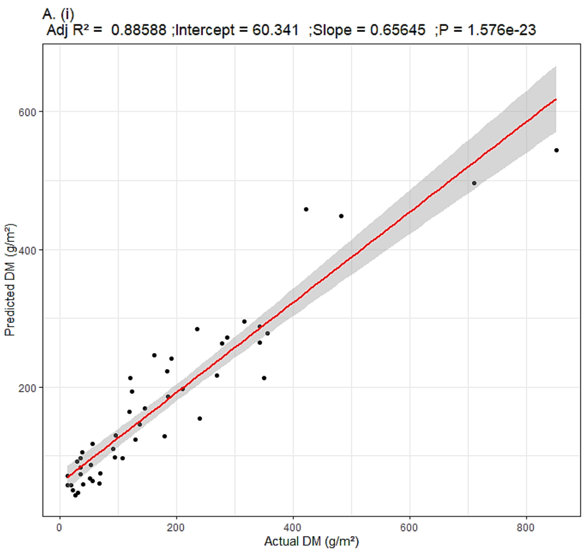


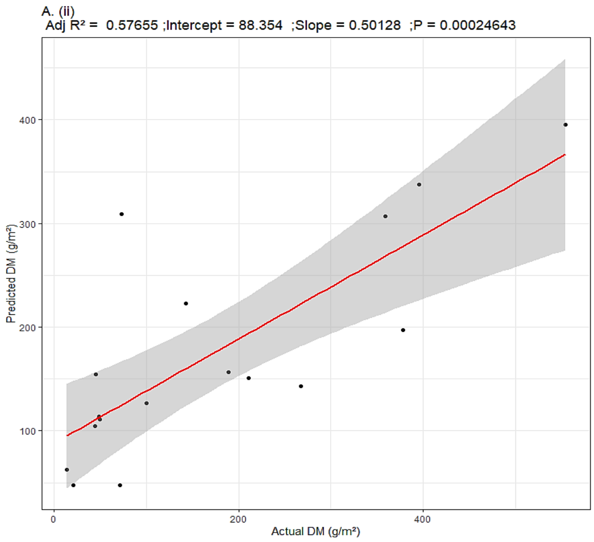

A. (i)
Adj R² 0.90898 ;Intercept = 1.4337 ;Slope = 0.78393 ;P = 8.5731e-26
Predicted CP (%)
Actual CP (%)

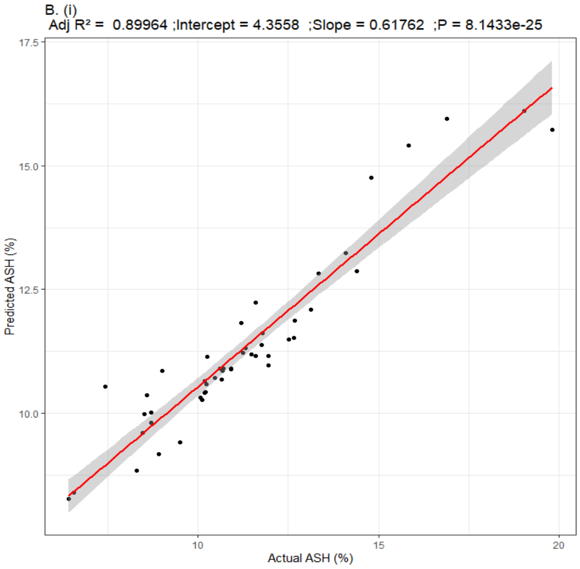

A. (ii)
Adj R² 0.40396 ;Intercept = 3.4112 ;Slope = 0.48641 ;P = 0.0036347
Predicted CP (%)
Actual CP (%)

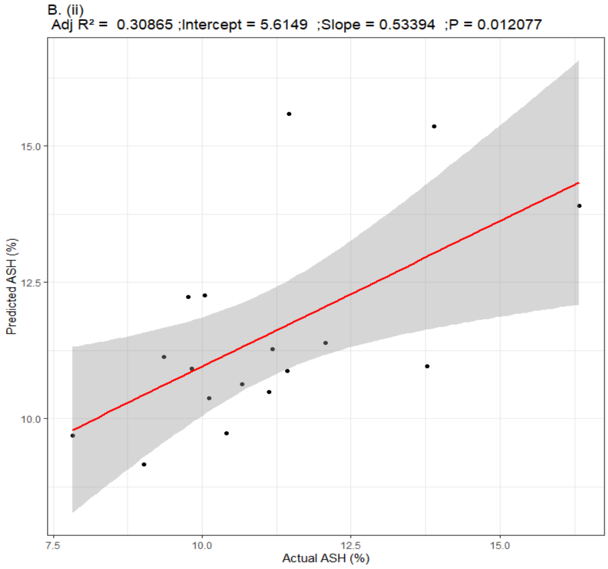

A. (i)
Adj R² 0.90336 ;Intercept = 11.58 ;Slope = 0.77861 ;P = 3.4082e-25
Predicted IVDMD (%)
Actual IVDMD (%)

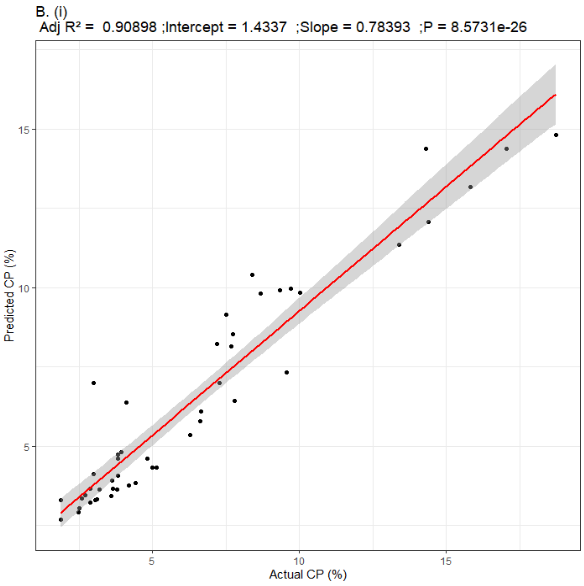

A. (ii)
Adj R² 0.4171 ;Intercept = 28.878 ;Slope = 0.48667 ;P = 0.0030417
Predicted IVDMD (%)
Actual IVDMD (%)

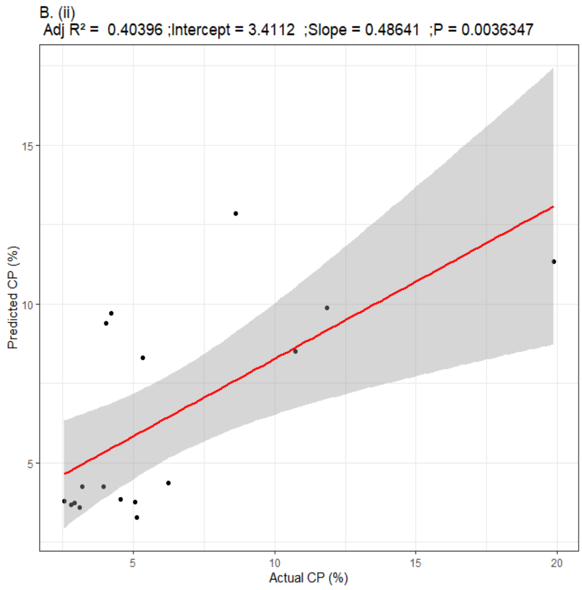

A. (i)
Adj R² = 0.88588 ;Intercept = 60.341 ;Slope = 0.65645 ;P = 1.576e-23
Predicted DM (g/m²)
Actual DM (g/m²)

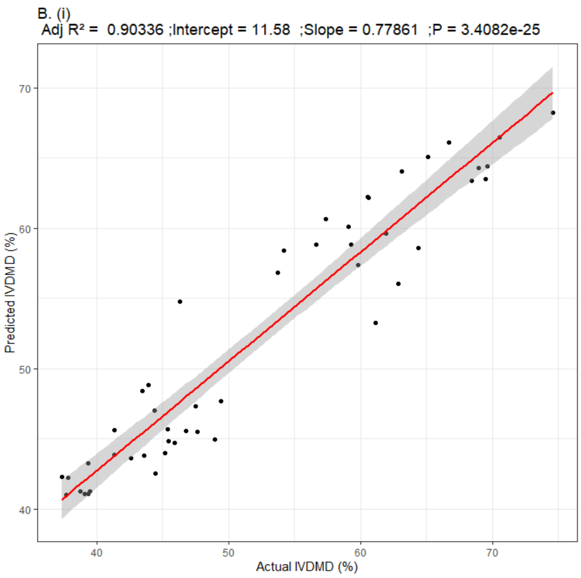

A. (ii)
Adj R² = 0.57655 ;Intercept = 88.354 ;Slope = 0.50128 ;P = 0.00024643
Predicted DM (g/m²)
Actual DM (g/m²)

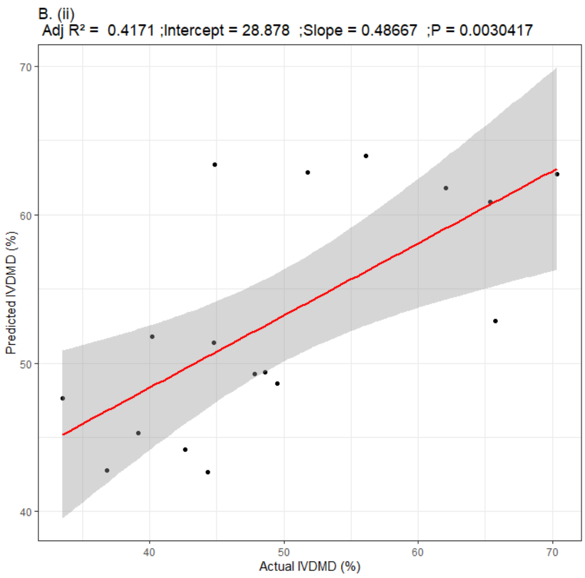

B. (i)
Adj R² = 0.89964 ;Intercept = 4.3558 ;Slope = 0.61762 ;P = 8.1433e-25
Predicted ASH (%)
Actual ASH (%)


B. (ii)
Adj R² = 0.30865 ;Intercept = 5.6149 ;Slope = 0.53394 ;P = 0.012077
Predicted ASH (%)
Actual ASH (%)


B. (i)
Adj R² = 0.90898 ;Intercept = 1.4337 ;Slope = 0.78393 ;P = 8.5731e-26
Predicted CP (%)
Actual CP (%)


B. (ii)
Adj R² = 0.40396 ;Intercept = 3.4112 ;Slope = 0.48641 ;P = 0.0036347
Predicted CP (%)
Actual CP (%)


B. (i)
Adj R² = 0.90336 ;Intercept = 11.58 ;Slope = 0.77861 ;P = 3.4082e-25
Predicted IVDMD (%)
Actual IVDMD (%)


B. (ii)
Adj R² = 0.4171 ;Intercept = 28.878 ;Slope = 0.48667 ;P = 0.0030417
Predicted IVDMD (%)
Actual IVDMD (%)

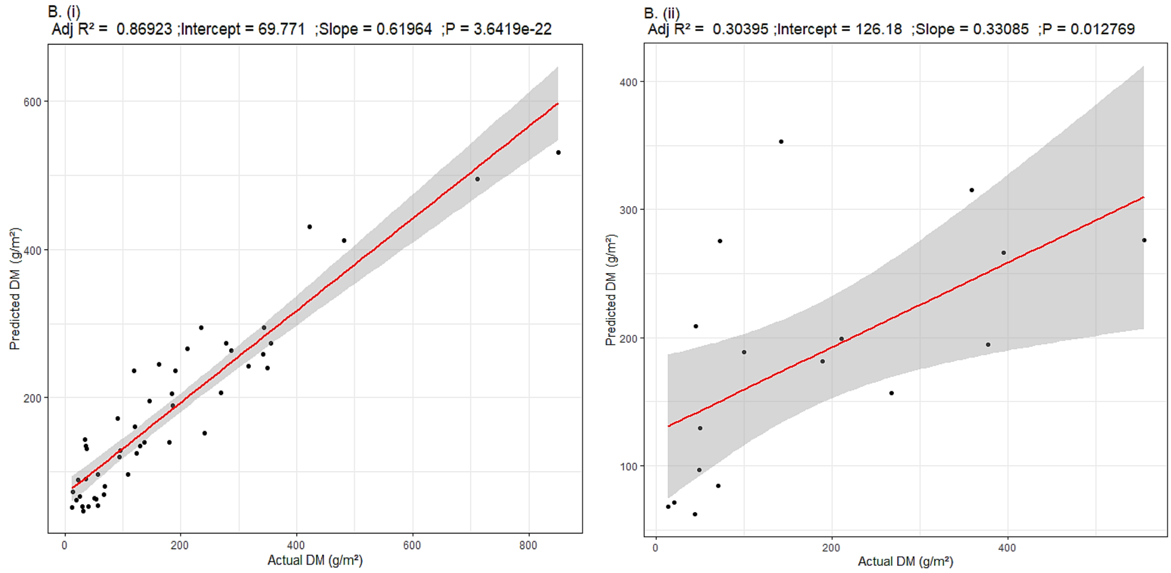


Figure 5.1. Relationship between actual and predicted Ash, CP, IVDMD and DM for (i) training (n= 48) and (ii) testing (n= 17) analyses using random forest regression model. The regression model was developed using A) VI+ Bands combination as it gave the optimum results B) only SR as it gave the worst result.

The variable importance showed handheld NDVI to be highly correlated with the given forage parameters hence it was excluded as it could be easily justified. Variables like the grass height, fresh and dry weight that could be highly related to the dry matter yields were not used in the predictor dataset as they had strong linear relationship.

5.3 Limitations

The test data shows low correlation as compared to the training dataset because of the low number of samples i.e. 17; the training data (48 samples) on the other hand shows high correlation with the predicted dataset as shown in figure 5.1. The training data performed reasonably well for estimating the forage quality and quantity parameters compared to testing data. The variable importance using mlr libraries caused irrelevant variables to be the most dominant, like Red band was shown to be the most important predictor for DM. Using the library caret gave more sensible important variables as shown in Figure 4.3. Random forest was used because it deals with small and large sample sizes with low bias, even with a small sample size the results were satisfactory, the training data was cross validated with 100 repetitions which resulted in an optimised and fairly accurate dataset.

The forage parameters oADF (acid detergent fibre) and oNDF (neutral detergent fibre) were not used in the predictor dataset as they had missing values for five samples. Cutler, (2010) suggested that missing data imputation for RF algorithm should be done using proximities although the faster way is by using the median of the non-missing values and replacing it with the missing values. This method of using proximities was not followed in this study as the parameters were excluded from the dataset, but it would be a suggestion for studies to try in the future.

6. Conclusion

A total of seven combinations of predictor variables was tested and VI + Bands was found to be the one with lowest RMSE among all which was considered to have the optimum values. Vegetation indices containing the only the Red and NIR bands were the most important variables, for example, DVI, SAVI, RVI for IVDMD, DM and CP respectively. As stated above, the red region is related with ash and protein concentration, hence the most important variable for Ash was the red band. The SR combination showed the lowest performance with the highest RMSE for all forage parameters; blue and green bands were rated the lowest as they rarely contributed to the validation dataset.

Using R as the statistical programming language to run the random forest regression algorithm was beneficial for the project as it contains in built packages to get an accurate result. Using both mlr and caret packages helped in the better understanding of how the code was structured around the dataset, although caret was used for cross validation and permutation. A custom RF model was made, with mtry and ntree values set manually according to the requirement of the dataset. Due to a smaller number of samples present, it was beneficial to set the tuning parameters manually as they were adjusted according to the sample data.

The spatial prediction was done taking the VI+ SR+ Bands combination into consideration. As the most important predictor variables for this combination were Red for Ash, NLI for CP and IVDMD and MSR for DM, the prediction was done using these particular variables. Although red band was rated the most important for Ash, the areas in the spatially predicted map were not distinct, it appeared smoothened out. The other three parameters were visualised clearly and showed the range of the important variables around the sampling location (Atoca in this case). This is a good practice to follow as it helps in assessing the whole area by predicting its values, rather than physically being there and collect the sampling points. This is why remote sensing plays a vital part in reducing the cost of sample collection and determining the areas which need further management.

To summarise, the present study has evaluated the possibilities of using Planet Scope data for estimating forage quantity and quality. Predictor variables derived from the Planet Scope sensors were enough to accurately predict Ash concentration and standing biomass dry weight (DM) from field observations. For crude protein and in vitro dry matter digestibility the results were less accurate. A repeated reduction of the predictor variable set was implemented, guided by a permutation-based variable importance measure. Thus, a subset of important variables was identified. The vegetation indices with NIR and Red bands in the combination of VI+ Bands were found to be particularly important. The future studies using the Planet Scope satellite dataset will hopefully benefit from this study and utilise this information.

References

Abdel-Rahman, E. M., Ahmed, F. B., & Ismail, R. (2013). Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data. *International Journal of Remote Sensing*, *34*(2), 712–728. https://doi.org/10.1080/01431161.2012.713142

Adelabu, S., Mutanga, O., & Adam, E. (2014). Evaluating the impact of red-edge band from Rapideye image for classifying insect defoliation levels. *ISPRS Journal of Photogrammetry and Remote Sensing*, *95*, 34–41. https://doi.org/10.1016/j.isprsjprs.2014.05.013

Ali, ;, Cawkwell, F. ;, Dwyer, E. ;, Green, S., Ali, I., Cawkwell, F., & Dwyer, E. (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 1 Modeling Managed Grassland Biomass Estimation by Using . *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), 3254–3264. https://doi.org/10.1109/JSTARS.2016.2561618

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., Green, S., & Dwyer, N. (2014). Application of statistical and machine learning models for grassland yield estimation based on a hypertemporal satellite remote sensing time series. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 5060–5063. https://doi.org/10.1109/IGARSS.2014.6947634

Asner, G. P., Martin, R. E., & Mascaro, J. (2017). Coral reef atoll assessment in the South China Sea using Planet Dove satellites. *Remote Sensing in Ecology and Conservation*, *3*(2), 57–65. https://doi.org/10.1002/rse2.42

Astola, H., Häme, T., Sirro, L., Molinier, M., & Kilpi, J. (2019). Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sensing of Environment*, *223*, 257–273. https://doi.org/10.1016/j.rse.2019.01.019

Auret, L., & Aldrich, C. (2012). Interpretation of nonlinear relationships between process variables by use of random forests. *Minerals Engineering*, *35*, 27–42. https://doi.org/10.1016/j.mineng.2012.05.008

Beckschäfer, P., Fehrmann, L., Harrison, R. D., Xu, J., & Kleinn, C. (2014). Mapping leaf area index in subtropical upland ecosystems using rapideye imagery and the randomforest algorithm. *IForest*, *7*(1), 1–11. https://doi.org/10.3832/ifor0968-006

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. In *ISPRS Journal of Photogrammetry and Remote Sensing* (Vol. 114, pp. 24–31). Elsevier B.V. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Berhane, T. M., Lane, C. R., Wu, Q., Autrey, B. C., Anenkhonov, O. A., Chepinoga, V. v., & Liu, H. (2018). Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. *Remote Sensing*, *10*(4), 580. https://doi.org/10.3390/rs10040580

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

Boddey, R. M., MacEdo, R., Tarré, R. M., Ferreira, E., de Oliveira, O. C., De, C., Cantarutti, R. B., Pereira, J. M., Alves, B. J. R., & Urquiaga, S. (2004). Nitrogen cycling in Brachiaria pastures: The key to understanding the process of pasture decline. *Agriculture, Ecosystems and Environment*, *103*(2), 389–403. https://doi.org/10.1016/j.agee.2003.12.010

Boegh, E., Soegaard, H., Broge, N., Schelde, K., Thomsen, A., Hasager, C. B., & Jensen, N. O. (2002). Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sensing of Environment*, *81*(2–3), 179–193. https://doi.org/10.1016/S0034-4257(01)00342-X

Borowik, T., Pettorelli, N., Sönnichsen, L., & Jędrzejewska, B. (2013). Normalized difference vegetation index (NDVI) as a predictor of forage availability for ungulates in forest and field habitats. *European Journal of Wildlife Research*, *59*(5), 675–682. https://doi.org/10.1007/s10344-013-0720-0

Boschetti, M., Bocchi, S., & Brivio, P. A. (2006). *Assessment of pasture production in the Italian Alps using spectrometric and remote sensing information*. https://doi.org/10.1016/j.agee.2006.05.024

Breiman, L. (2001). *Random Forests* (Vol. 45).

Breunig, F. M., Galvão, L. S., Dalagnol, R., Dauve, C. E., Parraga, A., Santi, A. L., della Flora, D. P., & Chen, S. (2020). Delineation of management zones in agricultural fields using cover–crop biomass estimates from PlanetScope data. *International Journal of Applied*

Earth Observation and Geoinformation, *85*, 102004. https://doi.org/10.1016/j.jag.2019.102004

Broadband Greenness. (2020). https://www.harrisgeospatial.com/docs/BroadbandGreenness.html

Brown, G. (2010). *Ensemble Learning Motivation and Background*. Springer Press.

Capolupo, A., Kooistra, L., Berendonk, C., Boccia, L., & Suomalainen, J. (2015). Estimating Plant Traits of Grasslands from UAV-Acquired Hyperspectral Images: A Comparison of Statistical Approaches. *ISPRS International Journal of Geo-Information*, *4*(4), 2792–2820. https://doi.org/10.3390/ijgi4042792

Capstaff, N. M., & Miller, A. J. (2018). Improving the yield and nutritional quality of forage crops. In *Frontiers in Plant Science* (Vol. 9, p. 535). Frontiers Media S.A. https://doi.org/10.3389/fpls.2018.00535

Chan, J. C. W., Beckers, P., Spanhove, T., & Borre, J. vanden. (2012). An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery. *International Journal of Applied Earth Observation and Geoinformation*, *18*(1), 13–22. https://doi.org/10.1016/j.jag.2012.01.002

Chen, J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, *22*(3), 229–242. https://doi.org/10.1080/07038992.1996.10855178

Cheng, Y. (2019). *RETRIEVING VEGETATION PHENOLOGY WITH PLANETSCOPE IMAGES FOR A SEMI-ARID RANGELAND IN KENYA*.

Clevers, J G P W, & Gitelson, A. A. (2012). *USING THE RED-EDGE BANDS ON SENTINEL-2 FOR RETRIEVING CANOPY CHLOROPHYLL AND NITROGEN CONTENT*.

Clevers, Jan G.P.W., Kooistra, L., & van den Brande, M. M. M. (2017). Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sensing*, *9*(5). https://doi.org/10.3390/rs9050405

Collins, M., Lacefield, G., Martin, N., Mertens, D., Olson, K., Putnam, D., Undersander, D., & Wolf, M. (n.d.). *Understanding forage quality Don Ball*.

Cox, S. (2002). Information technology: The global key to precision agriculture and sustainability. *Computers and Electronics in Agriculture*, *36*(2–3), 93–111. https://doi.org/10.1016/S0168-1699(02)00095-9

Crippen, R. E. (1990). Calculating the vegetation index faster. *Remote Sensing of Environment*, *34*(1), 71–73. https://doi.org/10.1016/0034-4257(90)90085-Z

Cutler, A. (2010). *Random Forests for Regression and Classification*.

Dangal, S. R. S., Tian, H., Lu, C., Pan, S., Pederson, N., & Hessl, A. (2016). Synergistic effects of climate change and grazing on net primary production of Mongolian grasslands. *Ecosphere*, *7*(5). https://doi.org/10.1002/ecs2.1274

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., & Cho, M. (2008). LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(4), 409–426. https://doi.org/10.1016/j.isprsjprs.2008.01.001

Departamento de Valle del Cauca (Colombia) - EcuRed. (2020). https://www.ecured.cu/Departamento_de_Valle_del_Cauca_(Colombia)

Department of Energy & Climate Change. (2020). *Business Case and Intervention Summary - Silvopastoral systems for climate change mitigation and poverty alleviation in Colombia´s livestock sector*.

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), 2534. https://doi.org/10.3390/rs12162534

Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, *118*, 259–272. https://doi.org/10.1016/j.rse.2011.11.020

Fact sheet - Dichanthium aristatum. (2020). http://www.tropicalforages.info/key/forages/Media/Html/entities/dichanthium_aristatum.htm

Felton, A. M., Wam, H. K., Stolter, C., Mathisen, K. M., & Wallgren, M. (2018). The complexity of interacting nutritional drivers behind food selection, a review of northern cervids. *Ecosphere*, *9*(5), e02230. https://doi.org/10.1002/ecs2.2230

Foody, G. M., Mathur, A., Sanchez-Hernandez, C., & Boyd, D. S. (2006). Training set size requirements for the classification of a specific class. *Remote Sensing of Environment, 104*(1), 1–14. https://doi.org/10.1016/j.rse.2006.03.004

Forage Quality Parameters Explained Agronomy Fact Sheet Series. (2016). http://ccedelaware.org/wp-

Franklin, D. R., & Martin, W. (1993). *F O R A G E S*. http://www.echonet.org/

Fulgueira, C. L., Amigot, S. L., Gaggiotti, M., Romero, L. A., & Basílico, J. C. (2007). *Forage Quality : Techniques for Testing*.

Gáborčík, N. (2003). Relationship between contents of chlorophyll (a+b) (SPAD values) and nitrogen of some temperate grasses. In *PHOTOSYNTHETICA* (Vol. 41, Issue 2).

Gašparović, M., Medak, D., Pilaš, I., Jurjević, L., & Balenović, I. (2018). *FUSION OF SENTINEL-2 AND PLANETSCOPE IMAGERY FOR VEGETATION DETECTION AND MONITORING*. https://doi.org/10.5194/isprs-archives-XLII-1-155-2018

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Gitelson, A. A., & Merzlyak, M. N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research*, *22*(5), 689–692. https://doi.org/10.1016/S0273-1177(97)01133-2

Gitelson1, A. A., Stark2, R., Grits2, U., Rundquist1, D., Kaufman3, Y., & Derry1, D. (2002). Vegetation and soil lines in visible spectral space: a concept and technique for remote estimation of vegetation fraction. *Int. j. Remote Sensing*, *23*(13), 2537–2562. https://doi.org/10.1080/01431160110107806

Goel, N. S., & Qin, W. (1994). Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: a computer simulation. *Remote Sensing Reviews*, *10*(4), 309–347. https://doi.org/10.1080/02757259409532252

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2017). Correlation and variable importance in random forests. *Statistics and Computing*, *27*(3), 659–678. https://doi.org/10.1007/s11222-016-9646-1

Grömping, U. (2015). *Variable importance in regression models*. https://doi.org/10.1002/wics.1346

Guo, X., Wilmshurst, J. F., & Li, Z. (2010). Comparison of laboratory and field remote sensing methods to measure forage quality. *International Journal of Environmental Research and Public Health*, *7*(9), 3513–3530. https://doi.org/10.3390/ijerph7093513

Ham, J. S., Chen, Y., Crawford, M. M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *43*(3), 492–501. https://doi.org/10.1109/TGRS.2004.842481

Helman, D., Bahat, I., Netzer, Y., Ben-Gal, A., Alchanatis, V., Peeters, A., & Cohen, Y. (2018). Using Time Series of High-Resolution Planet Satellite Images to Monitor Grapevine Stem Water Potential in Commercial Vineyards. *Remote Sensing*, *10*(10), 1615. https://doi.org/10.3390/rs10101615

Hijmans, R. J. (2019). Package 'raster' - Geographic Data Analysis and Modeling. In *CRAN Repository* (p. 250). https://cran.r-project.org/package=raster

Hoa, N. H. (2017). COMPARISON OF VARIOUS SPECTRAL INDICES FOR ESTIMATING MANGROVE COVERS USING PLANETSCOPE DATA: A CASE STUDY IN XUAN THUY NATIONAL PARK, NAM DINH PROVINCE. In *Management of Forest Resources and Environment JOURNAL OF FORESTRY SCIENCE AND TECHNOLOGY*. https://www.planet.com/explorer

Hooftman, D. A. P., & Bullock, J. M. (2012). Mapping to inform conservation: A case study of changes in semi-natural habitats and their connectivity over 70years. *Biological Conservation*, *145*(1), 30–38. https://doi.org/10.1016/j.biocon.2011.09.015

Houborg, R., & McCabe, M. F. (2018). A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sensing of Environment*, *209*, 211–226. https://doi.org/10.1016/j.rse.2018.02.067

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, *233*, 111410. https://doi.org/10.1016/j.rse.2019.111410

Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, *4*(9), 2661–2693. https://doi.org/10.3390/rs4092661

Ismail, R., & Mutanga, O. (2010). A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, *12*(SUPPL. 1), S45–S51. https://doi.org/10.1016/j.jag.2009.09.004

Isselstein, J., Jeangros, B., & Pavlu, V. (2005). Agronomic aspects of biodiversity targeted management of temperate grasslands in Europe-A review. In *Agronomy Research* (Vol. 3, Issue 2).

Kawamura, K., Mackay, A. D., Tuohy, M. P., Betteridge, K., Sanches, I. D., & Inoue, Y. (2011). Potential for spectral indices to remotely sense phosphorus and potassium content of legume-based pasture as a means of assessing soil phosphorus and potassium fertility status. *International Journal of Remote Sensing*, *32*(1), 103–124. https://doi.org/10.1080/01431160903439908

Kemp, D. R., & Michalk, D. L. (2007). Towards sustainable grassland and livestock management. In *Journal of Agricultural Science* (Vol. 145, Issue 6, pp. 543–564). https://doi.org/10.1017/S0021859607007253

Koppe, W., & li, fei. (2010). *Evaluating Multispectral and Hyperspectral Satellite Remote Sensing Data for Estimating Winter Wheat Growth Parameters at Regional Scale in the North China Plain*. https://doi.org/10.1127/1432-8364/2010/00471

Kuhn Max. (2020). *Package "caret" Title Classification and Regression Training*.

Lamarque, P., Lavorel, S., Mouchet, M., & Quétier, F. (2014). Plant trait-based models identify direct and indirect effects of climate change on bundles of grassland ecosystem services. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13751–13756. https://doi.org/10.1073/pnas.1216051111

Landsat 8 « Landsat Science. (2020). https://landsat.gsfc.nasa.gov/landsat-data-continuity-mission/

Li, F., Mistele, B., Hu, Y., Chen, X., & Schmidhalter, U. (2014). Reflectance estimation of canopy nitrogen content in winter wheat using optimised hyperspectral spectral indices and partial least squares regression. *European Journal of Agronomy*, *52*, 198–209. https://doi.org/10.1016/j.eja.2013.09.006

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. In *Sensors (Switzerland)* (Vol. 18, Issue 8). MDPI AG. https://doi.org/10.3390/s18082674

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest* (Vol. 2, Issue 3). http://www.stat.berkeley.edu/

Lima, A. R., Cannon, A. J., & Hsieh, W. W. (2015). Nonlinear regression in environmental sciences using extreme learning machines: A comparative evaluation. *Environmental Modelling and Software*, *73*, 175–188. https://doi.org/10.1016/j.envsoft.2015.08.002

Liu, H., Dahlgren, R. A., Larsen, R. E., Devine, S. M., Roche, L. M., O' Geen, A. T., Wong, A. J. Y., Covello, S., & Jin, Y. (2019). Estimating rangeland forage production using remote sensing data from a Small Unmanned Aerial System (sUAS) and planetscope satellite. *Remote Sensing*, *11*(5). https://doi.org/10.3390/rs11050595

Loch, D. S. (1977). Brachiaria decumbens (signal grass) - a review with particular reference to Australia. *Tropical Grasslands*, *11*(2), 141–157. https://www.feedipedia.org/node/21352

Loozen, Y., Karssenberg, D., de Jong, S. M., Wang, S., van Dijk, J., Wassen, M. J., & Rebel, K. T. (2019). Exploring the use of vegetation indices to sense canopy nitrogen to phosphorous ratio in grasses. *International Journal of Applied Earth Observation and Geoinformation*, *75*, 1–14. https://doi.org/10.1016/j.jag.2018.08.012

Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially located platform and aerial photography for documentation of grazing impacts on wheat. *Geocarto International*, *16*(1), 65–70. https://doi.org/10.1080/10106040108542184

Lu, B., He, Y., & Dao, P. D. (2019). Comparing the performance of multispectral and hyperspectral images for estimating vegetation properties. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(6), 1784–1797. https://doi.org/10.1109/JSTARS.2019.2910558

Mauya, E. W., Koskinen, J., Tegel, K., Hämäläinen, J., Kauranne, T., & Käyhkö, N. (2019). Modelling and Predicting the Growing Stock Volume in Small-Scale Plantation Forests of Tanzania Using Multi-Sensor Image Synergy. *Forests*, *10*(3), 279. https://doi.org/10.3390/f10030279

McNairn, H., Kross, A., Lapen, D., Caves, R., & Shang, J. (2014). Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. *International Journal of Applied Earth Observation and Geoinformation*, *28*(1), 252–259. https://doi.org/10.1016/j.jag.2013.12.015

Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(8), 1778–1790. https://doi.org/10.1109/TGRS.2004.831865

Merkle, J. A., Monteith, K. L., Aikens, E. O., Hayes, M. M., Hersey, K. R., Middleton, A. D., Oates, B. A., Sawyer, H., Scurlock, B. M., & Kauffman, M. J. (2016). Large herbivores surf waves of green-up during spring. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1833), 20160456. https://doi.org/10.1098/rspb.2016.0456

Meyer, H., Lehnert, L. W., Wang, Y., Reudenbach, C., Nauss, T., & Bendix, J. (2017). From local spectral measurements to maps of vegetation cover and biomass on the Qinghai-Tibet-Plateau: Do we need hyperspectral information? *International Journal of Applied Earth Observation and Geoinformation*, *55*, 21–31. https://doi.org/10.1016/j.jag.2016.10.001

Miller, G. J., Morris, J. T., & Wang, C. (2019). Estimating Aboveground Biomass and Its Spatial Distribution in Coastal Wetlands Utilizing Planet Multispectral Imagery. *Remote Sensing*, *11*(17), 2020. https://doi.org/10.3390/rs11172020

Mudereri, B. T., Dube, T., Adel-Rahman, E. M., Niassy, S., Kimathi, E., Khan, Z., & Landmann, T. (2019). *A comparative analysis of PlanetScope and Sentinel-2 space-borne sensors in mapping Striga weed using Guided Regularised Random Forest classification ensemble*. https://doi.org/10.5194/isprs-archives-XLII-2-W13-701-2019

Mutanga, O., Skidmore, A. K., Kumar, L., & Ferwerda, J. (2005). Estimating tropical pasture quality at canopy level using band depth analysis with continuum removal in the visible domain. *International Journal of Remote Sensing*, *26*(6), 1093–1108. https://doi.org/10.1080/01431160512331326738

Mutanga, O, Skidmore, A. K., & Prins, H. H. T. (n.d.). *Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features*. https://doi.org/10.1016/j.rse.2003.11.001

Mutanga, Onisimo, Adam, E., & Cho, M. A. (2012). High density biomass estimation for wetland vegetation using worldview-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation*, *18*(1), 399–406. https://doi.org/10.1016/j.jag.2012.03.012

Mutanga, Onisimo, & Skidmore, A. K. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *62*(1), 34–42. https://doi.org/10.1016/j.isprsjprs.2007.02.001

Otgonbayar, M., Atzberger, C., Chambers, J., & Damdinsuren, A. (2019). Mapping pasture biomass in Mongolia using Partial Least Squares, Random Forest regression and Landsat 8 imagery. *International Journal of Remote Sensing*, *40*(8), 3204–3226. https://doi.org/10.1080/01431161.2018.1541110

Palmer, S. C. F., Gordon, I. J., Hester, A. J., & Pakeman, R. J. (2005). Introducing spatial grazing impacts into the prediction of moorland vegetation dynamics. *Landscape Ecology*, *19*(8), 817–827. https://doi.org/10.1007/s10980-005-0094-0

Parente, L., Taquary, E., Silva, A., Souza, C., & Ferreira, L. (2019). Next Generation Mapping: Combining Deep Learning, Cloud Computing, and Big Remote Sensing Data. *Remote Sensing*, *11*(23), 2881. https://doi.org/10.3390/rs11232881

Prasad, A. M., Iverson, L. R., & Liaw, A. (2006). Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, *9*(2), 181–199. https://doi.org/10.1007/s10021-005-0054-1

Prepared for the Plant Production and Protection Division Food and Agriculture Organization of the United Nations (FAO) Challenges and opportunities for carbon sequestration in grassland systems A technical report on grassland management and climate change mitigation. (2010).

Punalekar, S. M., Verhoef, A., Quaife, T. L., Humphries, D., Bermingham, L., & Reynolds, C. K. (2018). Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sensing of Environment*, *218*, 207–220. https://doi.org/10.1016/j.rse.2018.09.028

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). *A Modified Soil Adjusted Vegetation Index* (Vol. 48).

R: A Language and Environment for Statistical Computing Reference Index The R Core Team. (2020). https://www.gnu.org/copyleft/gpl.html.

Raab, C., Riesch, F., Tonn, B., Barrett, B., Meißner, M., Balkenhol, N., & Isselstein, J. (2020). Target-oriented habitat and wildlife management: estimating forage quantity and quality of semi-natural grasslands with Sentinel-1 and Sentinel-2 data. *Remote Sensing in Ecology and Conservation*, rse2.149. https://doi.org/10.1002/rse2.149

Ramoelo, A., Cho, M. A., Mathieu, R., Madonsela, S., van de Kerchove, R., Kaszta, Z., & Wolff, E. (2015). Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and WorldView-2 data. *International Journal of*

Applied Earth Observation and Geoinformation, *43*, 43–54. https://doi.org/10.1016/j.jag.2014.12.010

Ramoelo, A., Cho, M., Mathieu, R., & Skidmore, A. K. (2015). Potential of Sentinel-2 spectral configuration to assess rangeland quality. *Journal of Applied Remote Sensing*, *9*(1), 094096. https://doi.org/10.1117/1.jrs.9.094096

Ramoelo, A., Cho, M., Mathieu, R., Skidmore, A., Schlerf, M., & Heitkönig, I. (2012). Estimating grass nutrients and biomass as an indicator of rangeland (forage) quality and quantity using remote sensing in Savanna ecosystems. *Researchspace*, 1–8. http://srtm.csi.cgiar.org/

Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*.

Raynor, E. J., Joern, A., Nippert, J. B., & Briggs, J. M. (2016). Foraging decisions underlying restricted space use: effects of fire and forage maturation on large herbivore nutrient uptake. *Ecology and Evolution*, *6*(16), 5843–5853. https://doi.org/10.1002/ece3.2304

Reddersen, B., Fricke, T., & Wachendorf, M. (2014). A multi-sensor approach for predicting biomass of extensively managed grassland. *Computers and Electronics in Agriculture*, *109*, 247–260. https://doi.org/10.1016/j.compag.2014.10.011

Rincón Castillo, Á., Álvarez León, M. de, Pardo Barbosa, O., Alejandra Amaya, M., & Raúl Alejandro Díaz Giraldo, Y. (2019). Estimating chlorophyll concentration and its relationship with crude protein concentration in three species of Urochloa in the Piedemonte Llanero, Colombia. *Tropical Grasslands-Forrajes Tropicales*, *7*(5), 533–537. https://doi.org/10.17138/TGFT(7)533-537

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*(1), 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

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

Rouse, J. W. ;, Haas, R. H. ;, Schell, J. A. ;, & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *Proceedings of the Third Earth Resources Technology Satellite- 1 Symposium*, 301–317.

Saini, R., & Ghosh, S. K. (2017). Ensemble classifiers in remote sensing: A review. *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2017*, *2017-January*, 1149–1152. https://doi.org/10.1109/CCAA.2017.8229969

Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment*, *95*(2), 177–194. https://doi.org/10.1016/j.rse.2004.12.016

Schuster, C., Ali, I., Lohmann, P., Frick, A., Förster, M., & Kleinschmit, B. (2011). Remote Sensing Towards Detecting Swath Events in TerraSAR-X Time Series to Establish NATURA 2000 Grassland Habitat Swath Management as Monitoring Parameter. *Remote Sens, 3,* 1308–1322. https://doi.org/10.3390/rs3071308

Sentinel-2 - Missions - Sentinel Online. (2020). https://sentinel.esa.int/web/sentinel/missions/sentinel-2

Shimizu, K., Ota, T., Mizoue, N., & Saito, H. (2020). Comparison of Multi-Temporal PlanetScope Data with Landsat 8 and Sentinel-2 Data for Estimating Airborne LiDAR Derived Canopy Height in Temperate Forests. *Remote Sensing*, *12*(11), 1876. https://doi.org/10.3390/rs12111876

Shintate Galindo, F., Beloni, T., Buzetti, S., Carvalho, M., Filho, M. T., Dupas, E., Gaioto, M., & Ludkiewicz, Z. (2018). Technical and economic viability and nutritional quality of mombasa guinea grass silage production. *CROP PRODUCTION Acta Scientiarum. Agronomy*, *v*, *40*. https://doi.org/10.4025/actasciagron.v40i1.36395

Signal grass (Brachiaria decumbens) | Feedipedia. (2020). https://www.feedipedia.org/node/489

Sripada, Ravi P., Heiniger, R. W., White, J. G., & Meijer, A. D. (2006a). Aerial color infrared photography for determining early in-season nitrogen requirements in corn. *Agronomy Journal*, *98*(4), 968–977. https://doi.org/10.2134/agronj2005.0200

Sripada, Ravi P., Heiniger, R. W., White, J. G., & Meijer, A. D. (2006b). Aerial Color Infrared Photography for Determining Early In-Season Nitrogen Requirements in Corn. *Agronomy Journal*, *98*(4), 968–977. https://doi.org/10.2134/agronj2005.0200

Sripada, Ravi Prakash. (2005). *Determining In-Season Nitrogen Requirements for Corn Using Aerial Color-Infrared Photography*.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1), 307. https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*, 25. https://doi.org/10.1186/1471-2105-8-25

Suttie J.M, Reynolds S.G., & Battello C. (2005). *Grasslands of the world* (P. P. and P. S. Food and Agriculture Organization of the United Nations, Ed.). Food and Agriculture Organization.

Tälle, M. (2018). *Conservation of semi-natural grasslands: Effects of different management methods on biodiversity*.

The challenge of abandonment for the sustainable management of Palaearctic natural and semi-natural grasslands. (n.d.). https://doi.org/10.1515/hacq-2017-0018

Tian, Y., Yao, X., Yang, J., Cao, W., & Zhu, Y. (2011). *Plant Production Science Extracting Red Edge Position Parameters from Ground-and Space-Based Hyperspectral Data for Estimation of Canopy Leaf Nitrogen Concentration in Rice*. https://doi.org/10.1626/pps.14.270

Tong, A., & He, Y. (2017a). Estimating and mapping chlorophyll content for a heterogeneous grassland: Comparing prediction power of a suite of vegetation indices across scales between years. *ISPRS Journal of Photogrammetry and Remote Sensing*, *126*, 146–167. https://doi.org/10.1016/j.isprsjprs.2017.02.010

Tong, A., & He, Y. (2017b). Estimating and mapping chlorophyll content for a heterogeneous grassland: Comparing prediction power of a suite of vegetation indices across scales between years. *ISPRS Journal of Photogrammetry and Remote Sensing*, *126*, 146–167. https://doi.org/10.1016/j.isprsjprs.2017.02.010

Tucker, C. J. (1977). Spectral estimation of grass canopy variables. *Remote Sensing of Environment*, *6*(1), 11–26. https://doi.org/10.1016/0034-4257(77)90016-5

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Ustuner, M., Sanli, F. B., & Abdikan, S. (2016). *BALANCED VS IMBALANCED TRAINING DATA: CLASSIFYING RAPIDEYE DATA WITH SUPPORT VECTOR MACHINES*. https://doi.org/10.5194/isprsarchives-XLI-B7-379-2016

Varieties - Tropical Seeds. (2020). https://www.tropseeds.com/varieties/

Vescovo, L., Wohlfahrt, G., Balzarolo, M., Pilloni, S., Sottocornola, M., Rodeghiero, M., & Gianelle, D. (2012). New spectral vegetation indices based on the near-infrared shoulder wavelengths for remote detection of grassland phytomass. *International Journal of Remote Sensing*, *33*(7), 2178–2195. https://doi.org/10.1080/01431161.2011.607195

Vliegher, D. A., J-l, P., & den Pol -van Dasselaar, van A. (n.d.). Grassland term definitions and classifications adapted to the diversity of European grassland-based systems. In *Grassland Science in Europe* (Vol. 19).

Wachendorf, M. (2018). *Advances in remote sensing for monitoring grassland and forage production* (pp. 353–362). https://doi.org/10.19103/as.2017.0024.11

Waldén, E. (n.d.). *RESTORATION OF SEMI-NATURAL GRASSLANDS*.

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., Chang, Q. I., Yang, J., Zhang, X., & Li, F. (2018). *Estimation of paddy rice leaf area index using machine learning methods based on hyperspectral data from multi-year experiments*. https://doi.org/10.1371/journal.pone.0207624

Wei, C., Huang, J., Mansaray, L., Li, Z., Liu, W., & Han, J. (2017). Estimation and Mapping of Winter Oilseed Rape LAI from High Spatial Resolution Satellite Data Based on a Hybrid Method. *Remote Sensing*, *9*(5), 488. https://doi.org/10.3390/rs9050488

White, J. (2018). *Ash can trash your forage value*. https://todaysfarmermagazine.com/mag/livestock/1435-ash-can-trash-your-forage-value

Wijesingha, J., Astor, T., Schulze-Brüninghoff, D., Wengert, M., & Wachendorf, M. (2020). Predicting Forage Quality of Grasslands Using UAV-Borne Imaging Spectroscopy. *Remote Sensing*, *12*(1), 126. https://doi.org/10.3390/rs12010126

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1). https://doi.org/10.18637/jss.v077.i01

Xue, J., & Su, B. (2017). *Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications*. https://doi.org/10.1155/2017/1353691

Yadav, V. P., Prasad, R., Bala, R., Vishwakarma, A. K., Yadav, S. A., & Singh, S. K. (2019). *A COMPARISON OF MACHINE-LEARNING REGRESSION ALGORITHMS FOR THE*

ESTIMATION OF LAI USING LANDSAT-8 SATELLITE DATA. https://doi.org/10.5194/isprs-archives-XLII-4-W16-679-2019

Yuan, H., Yang, G., Li, C., Wang, Y., Liu, J., Yu, H., Feng, H., Xu, B., Zhao, X., & Yang, X. (2017). Retrieving Soybean Leaf Area Index from Unmanned Aerial Vehicle Hyperspectral Remote Sensing: Analysis of RF, ANN, and SVM Regression Models. *Remote Sensing*, *9*(4), 309. https://doi.org/10.3390/rs9040309

Zandler, H., Brenning, A., & Samimi, C. (2015). Quantifying dwarf shrub biomass in an arid environment: Comparing empirical methods in a high dimensional setting. *Remote Sensing of Environment*, *158*, 140–155. https://doi.org/10.1016/j.rse.2014.11.007

Zha, Y., Gao, J., Ni, S., Liu, Y., Jiang, J., & Wei, Y. (2003a). A spectral reflectance-based approach to quantification of grassland cover from Landsat TM imagery. *Remote Sensing of Environment*, *87*(2–3), 371–375. https://doi.org/10.1016/j.rse.2003.05.001

Zha, Y., Gao, J., Ni, S., Liu, Y., Jiang, J., & Wei, Y. (2003b). A spectral reflectance-based approach to quantification of grassland cover from Landsat TM imagery. *Remote Sensing of Environment*, *87*(2–3), 371–375. https://doi.org/10.1016/j.rse.2003.05.001

Zhang, B., Zhang, L., Xie, D., Yin, X., Liu, C., & Liu, G. (2015). Application of Synthetic NDVI Time Series Blended from Landsat and MODIS Data for Grassland Biomass Estimation. *Remote Sensing*, *8*(1), 10. https://doi.org/10.3390/rs8010010

Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. In *Precision Agriculture* (Vol. 13, Issue 6, pp. 693–712). Springer. https://doi.org/10.1007/s11119-012-9274-5

Zhang, H. K., & Roy, D. P. (2017). Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sensing of Environment*, *197*, 15–34. https://doi.org/10.1016/j.rse.2017.05.024

Zhen-wang, L., Xiao-ping, X., Huan, T., Fan, Y., Bao-rui, C., & Bao-hui, Z. (2017). ScienceDirect Estimating grassland LAI using the Random Forests approach and Landsat imagery in the meadow steppe of Hulunber, China. *Journal of Integrative Agriculture*, *2017*(2), 286–297. https://doi.org/10.1016/S2095-3119(15)61303-X

GES MSc Project Supervision Logbook

Student Name: Anushka Ghildiyal Supervisor(s): Dr Brian Barrett

Date	Type*	Duration	Topic/ Comments
05/05/2020	e-mail		Discussion on the chosen topic, details of the dataset and literature was suggested Received the datasets
09/06/2020	e-mail		YouTube tutorials were recommended to use R software
12/06/2020	e-mail		Received the PlanetScope satellite acquisitions, and associated ground truth and GPS data of the sampling points. Feedback on the research proposal received
16/06/2020	IM	30 minutes	Questions on how to get started with these datasets answered
21/06/2020	e-mail		Doubts regarding the survey data and how to extract the points on a map into a .csv file answered
22/06/2020	e-mail		Received details for the remote access to a dedicated university workstation due to problems with software installed on the laptop
15/07/2020	e-mail		Feedback on the codes used and plots. Solution provided to extract the raster values (of the sample locations) using a tool in ArcMap. Problems with the missing sample data discussed
20/07/2020	IM Chapter	30 minutes	Discussion on the problems faced with point extraction and RTK samples in the dataset. Feedback on the Chapter 1 received and negative R^2 values while doing variable optimisation discussed
06/07/2020	e-mail		Discussion with a PhD student in absence of the supervisor about the negative R^2 values and trying a different coding package (caret). Suggestions on which variables to use in the optimisation were helpful.
13/08/2020	e-mail		Errors with doing the spatial prediction discussed
18/07/2020	IM Skim	1 hour 30 minutes	Speed read of the project Formatting and grammatical errors were pointed out It was suggested to ask for an extension
19/07/2020	e-mail		Extension received

* Type of supervision/input

GM group meeting (in person or virtual)

IM individual meeting (in person or virtual)

Phone

e-mail correspondence.

Chapter record when you submitted a draft chapter for review

Skim record when you submit a draft report for skim reading.