Spatial Analysis of Habitat Quality in Semi-Natural Grasslands & Karst Landscapes Using Multi-Spectral Imagery: Case Study in Inis Mor, Ireland

Diploma Thesis

Author: Nicholas Marco Liquori

Supervisor: Ing. Vojtěch Barták, Ph.D.

2022

Faculty of Environmental Sciences

Department of Landscape and Urban Planning

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Environmental Sciences

DIPLOMA THESIS ASSIGNMENT

Nicholas Liquori, BA

Landscape Engineering
Landscape Planning

Thesis title

Spatial Analysis of Habitat Quality in Semi-Natural Grasslands & Karst Landscapes Using Multi-Spectral Imagery: Case Study in Inis Mor, Ireland

---

Objectives of thesis

The goal of this research is to validate the ability to assess habitat quality in semi- natural grasslands Caomhnú Árann using multi-spectral imagery. If results find that a statistical relationship exists between habitat quality and one or more spatial variables, then it can help identify the most relevant and sufficient variables for predicting priority habitat in pastoral semi-natural grasslands in Northern Europe. This will help in preliminary assessment of areas which may benefit from agri-environmental schemes like Caomhnú Árann and assist existing schemes to expand their project scale or lessen the labor expenditure on field scientists.

Methodology

A statistical model (random forest) will be used to test relationships between various spatial variables and the habitat quality of grazed grasslands in Inish Mor, Ireland. The statistical model will use remote sensing data from a multispectral drone survey conducted in June 2021 in attempt to find valid predictors of priority habitat (As defined in the "Grazing Scoring System" of the Agri-Env Scheme Caomhnú Árann), which will be ground truthed by an on-site assessment done in the summer of 2021. The accuracy of the random forest classification method will be cross referenced by the ground truthing data, and fields which have not been given a habitat score by Caomhnú Árann will be predicted by the model.

The proposed extent of the thesis

40 – 60 pages

Keywords

Multi-Spectral Imagery; Semi-Natural Grasslands; Vegetation Index; Agri-Environmental Schemes; Random Forest Classification

---

Recommended information sources

Jan Komárek, Tomáš Klouček, Jiří Prošek (2018) The potential of Unmanned Aerial Systems: A tool towards precision classification of hard-to-distinguish vegetation types? International Journal of Applied Earth Observation and Geoinformation, Volume 71, 2018, Pages 9-19, ISSN 0303-2434, https://doi.org/10.1016/j.jag.2018.05.003.

McGurn, P, Browne, A. , Ní Chonghaile, G., Duignan, L., Moran, J., ÓhUallacháin, D., Finn, J.A. "Semi-natural grasslands on the Aran Islands, Ireland: ecologically rich, economically poor". Farming for Conservation

Strong CJ, Burnside NG, Llewellyn D (2017) The potential of small-Unmanned Aircraft Systems for the rapid detection of threatened unimproved grassland communities using an Enhanced Normalized Difference Vegetation Index. PLoS ONE 12(10): e0186193. https://doi.org/10.1371/journal.pone.0186193

Yuhong He, Xulin Guo & John Wilmshurst (2006) Studying mixed grassland ecosystems I: suitable hyperspectral vegetation indices, Canadian Journal of Remote Sensing, 32:2, 98-107, DOI: 10.5589/m06-009

Expected date of thesis defence

2021/22 SS – FES

The Diploma Thesis Supervisor

Ing. Vojtěch Barták, Ph.D.

Supervising department

Department of Spatial Sciences

Electronic approval: 21. 3. 2022

doc. Ing. Petra Šímová, Ph.D.

Head of department

Electronic approval: 21. 3. 2022

prof. RNDr. Vladimír Bejček, CSc.

Dean

Prague on 31. 03. 2022

Author Acknowledgments

I would like to thank Ing. Vojtěch Barták, Ph.D. for his inspiration and guidance during the duration of this thesis research and I would like to thank Dr. Patrick McGurn and Dr. Amanda Browne from Caomhnú Árann for their assistance in this thesis research and the work they do on the Aran Islands that made this research possible.

I hereby declare that I have independently elaborated the diploma/final thesis with the topic of: Spatial Analysis of Habitat Quality in Semi-Natural Grasslands & Karst Landscapes Using Multi-Spectral Imagery: Case Study in Inis Mor, Ireland
and that I have cited all the information sources that I used in the thesis and that are also listed at the end of the thesis in the list of used information sources.
I am aware that my diploma/final thesis is subject to Act No. 121/2000 Coll., on copyright, on rights related to copyright and on amendment of some acts, as amended by later regulations, particularly the provisions of Section 35(3) of the act on the use of the thesis. I am aware that by submitting the diploma/final thesis I agree with its publication under Act No. 111/1998 Coll., on universities and on the change and amendments of some acts, as amended, regardless of the result of its defence.
With my own signature, I also declare that the electronic version is identical to the printed version and the data stated in the thesis has been processed in relation to the GDPR.

Nicholas Marco Liquori
Prague, CZ 31.31.2022

Abstract

Results Based Agri-Environmental Payment Schemes are powerful tools for protecting HNV farmland in Europe. The landscape present on the Aran Islands is one mosaiced by species-rich semi-natural grasslands and exposed limestone bedrock. The presence of rare flora species is owed to long-standing, traditional, low-intensity grazing practices. Caomhnú Árann monitors the farms of those who have enrolled in the program and provides a grade 1-5 ultimately depending on grazing level, presence of indicator species, adequate water provisions for cattle, and scrub maintenance. Unmanned Aerial Vehicle (UAV) acquired multispectral data can aid in the classification of these farmlands and limit the expenditure of time, money, and resources by Caomhnú Árann. Supervised classification by Random Forest Models was used to predict the grazing score of input fields with the highest overall accuracy by a single model of (54.78 %). This accuracy was achieved by the Random Forest Model which used the first 10 dimensions from a Principal Component Analysis as predictors. This model also achieved the highest balanced accuracy of any individual grazing class (class 5 – 84.01%). The variables which underwent the PCA and that built the other random forest models were various zonal statistics of the individual bands from the multispectral imagery, Normalized Difference Vegetation Index (NDVI) and NDVI derived analytical rasters, and a Digital Elevation Model (DEM). This research may have not successfully proven the full employability of a particular classification model as it pertains to the conservation interests of Caomhnú Árann and the Aran Islands, but there is promise in the ability to classify specific grazing scores. Scores 2 and 5 obtained the highest Balanced Accuracies across all 6 Random Forest Models. The results show that a Random Forest Model could be used by Caomhnú Árann in their future research if the collinearity of input data is corrected by performing a Principal Component Analysis, and input data is processed to limit the errors inherent with UAV acquired data.

Keywords: Multi-Spectral Imagery; Semi-Natural Grasslands; Vegetation Index; Agri-Environmental Schemes; Random Forest Classification

Abstract

Results Based Agri-Environmental Payment Schemes jsou účinnými nástroji pro ochranu zemědělské půdy s vysokou přírodní hodnotou v Evropě. Krajina Aranských ostrovů je tvořena mozaikou druhově bohatých polopřirozených pastvin a odkrytých vápencových skalních útvarů. Dlouhodobé, tradiční pastevní postupy s nízkou intenzitou uplatňované v této krajině vyústily v přítomnost vzácných druhů rostlin. Projekt Caomhnú Árann monitoruje farmy zapojené do programu a v závislosti na intenzitě pastvy, přítomnosti indikačních druhů, adekvátních zásobách vody pro dobytek a údržbě křovin jim přiděluje známku 1-5 (tzv. pastevní skóre). Multispektrální data získaná bezpilotními leteckými systémy (UAV) mohou pomoci při klasifikaci těchto zemědělských oblastí a omezit tak časové, finanční a jiné výdaje projektu Caomhnú Árann. V této práci byla k predikci pastevního skóre použita řízená klasifikace UAV snímků metodou Random Forest. Nejvyšší celkové přesnosti (54,78 %) bylo dosaženo v modelu, v němž bylo jako prediktorů použito prvních 10 hlavních komponent vypočtených ze všech dostupných vstupních vrstev. Tento model také dosáhl nejvyšší vyvážené přesnosti v rámci jednotlivých tříd pastvy (třída 5 – 84,01 %). Proměnné, které sloužily jako vstupy do analýzy hlavních komponent či jako samostatné prediktory v jednotlivých modelech, byly různé zonální statistiky jednotlivých pásem z multispektrálního snímku, normalizovaný vegetační index (NDVI), analytické rastry odvozené od NDVI a digitální model nadmořské výšky (DEM). Tato práce sice neprokázala plnou použitelnost konkrétního klasifikačního modelu pro účely ochrany přírody Aranských ostrovů v rámci projektu Caomhnú Árann, naznačuje však, že určitá konkrétní pastevní skóre klasifikovat lze. Skóre 2 a 5 získaly nejvyšší vyváženou přesnost (balanced accuracy) ze všech 6 modelů. Výsledky ukazují, že Random Forest modely by mohly být použity v dalším výzkumu v rámci projektu Caomhnú Árann, za předpokladu vyřešení kolinearity vstupních dat provedením analýzy hlavních komponent a předzpracování vstupních dat z UAV tak, aby byly opraveny chyby v těchto datech.

Table of Contents

Chapter 1

Introduction

1.1 Language Note

As the islands are a part of the Gaeltacht, this thesis will attempt to best retain the Irish spelling of place names, where it does not interfere with the ability of English Speakers to read along. Inis Meáin and Inis Oírr will retain their Irish spelling, but the largest island and that of our subject site, Árainn will be referred to as Inis Mór to not confuse English speakers with the name for the whole island chain "The Aran Islands". This is done out of respect for the community and individuals who made this effort possible. As *Tim Robinson* (2008) spiritually puts it in his book "Stones of Aran: Pilgrimage", "since the islands are a principal part of the Irish language's last precarious foothold on the world" an ethical duty is placed upon outsiders to respect the culture and traditions of the place they wish to observe science.

1.2 Introduction

Caomhnú Árann is a project under the EU's European Innovation Partnership for Agricultural productivity and Sustainability (EIP-AGRI) funded by the Department of Agriculture, Food and the Marine (DAFM). The project monitors 2,307 hectares of grazing land across all three islands within the Aran Islands (Caomhnú Árann 2020b). See Appendix A for the location of the Árann Islands and the subject site.

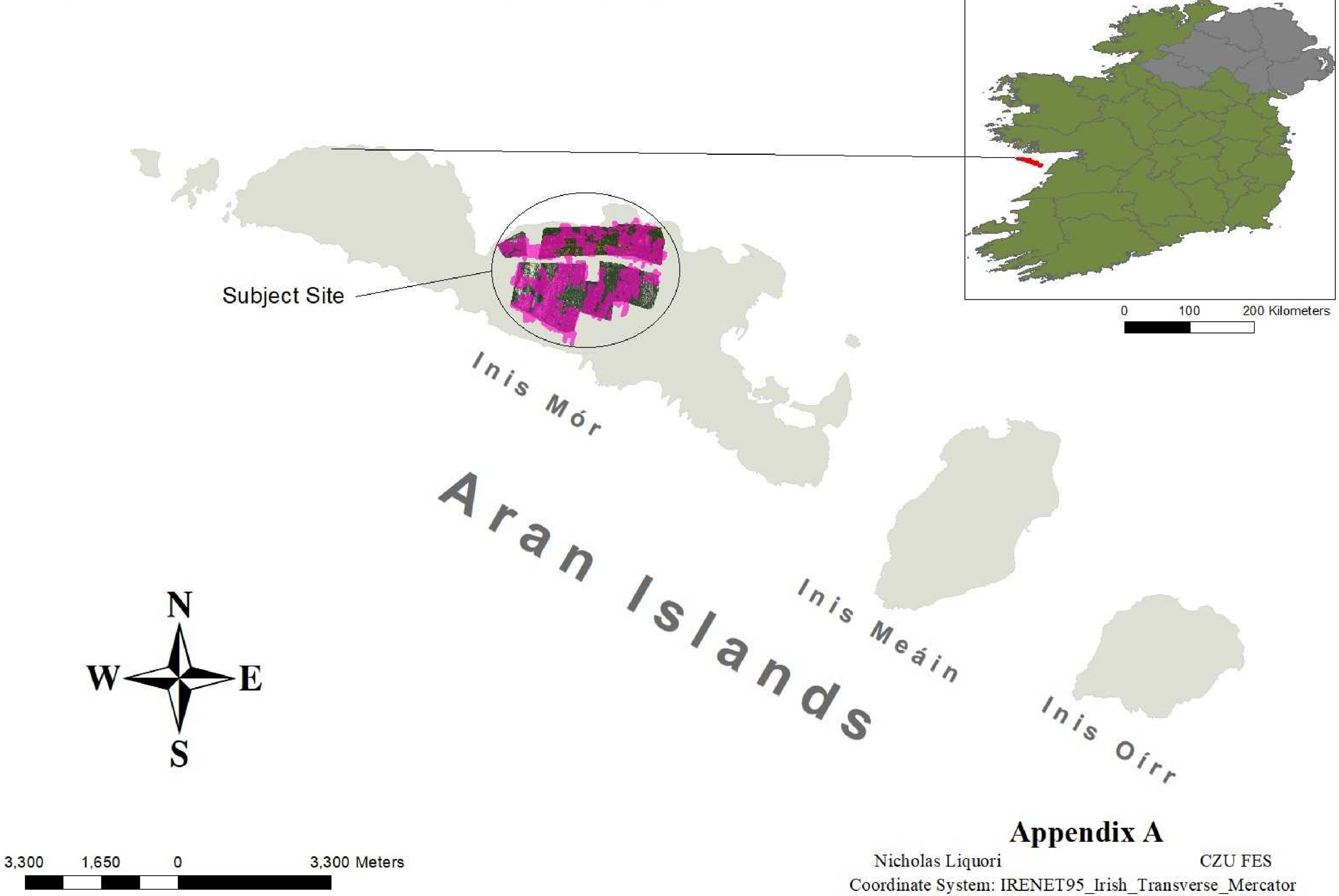
The project seeks to encourage farmers to maintain traditional, low- intensity grazing practices in an effort to conserve priority habitat. Modern socio-natural relations on the Aran Islands and The Burren have trended towards two farming outcomes: intensification or abandonment (McGurn 2017, McGurn 2020b). Factors of these relations include "low farm income, small and fragmented holdings, low productivity land, and high labor intensity of optimal conservation methods" (EIP-AGRI 2019). Farmers have been either abandoning their agricultural practices for jobs in the tourism sector and jobs off-island, or intensifying practices to increase income (Rensburg 2009). Both poles on this spectrum contribute to a decrease in biodiversity, and degradation of the natural heritage which drives the tourism sector.

Caomhnú Árann is a successor to the AranLIFE Project, which was funded by the EUs LIFE+ program and ran from 2014 to 2018. The project worked with farmers to assess whether they would be interested in participating in a results based payment scheme, and if such would positively contribute to species rich grassland habitats found on the islands. Aran Island farmers were exposed to the success of the BurrenLIFE project in the Burren, County Clare. The Burren is an eco-region hosting similar species rich grasslands and exposed limestone pavement, with a history of low-intensity grazing (McGurn 2020). BurrenLIFE would become a model for farmers on the Aran Islands looking to protect their natural heritage and economic livelihoods.

Caomhnú Árann developed a simple scoring system that grades for the quality of grassland habitat driven by grazing levels. Fields are scored one through five, with a score of five receiving the highest payments. Payments are also provided for active improvement measures of scrub control and installation of adequate water provisions. All three factors are key to the production of species-rich grasslands. The first requirement for a field's eligibility within Caomhnú Árann is that it must be grazed.

Due to the traveling limitations inherent with an island chain, it is in the interest of Caomhnú Árann to be able to score fields with as little ground-truthing as possible. An informational campaign was run during the tenure of AranLIFE, and now Caomhnú Árann, to make farmers knowledgeable on desirable grazing techniques and corrective action. A goal of this campaign is to enable farmers to grade farms on their own. In 2020, 25 farmers were randomly selected to grade their own fields using resources provided by Caomhnú Árann (Browne 2020). While additional training has been delayed due to travel restrictions brought upon by the Covid-19 pandemic, there exists a strong correlation between the scores issued by farmers and Caomhnú Árann. A combination of self-assessment done by farmers and verification by remote sensing will lessen project expenses and put those funds back into the pockets of farmers. The data coverage included in this thesis is limited to three small adjacent sites on Inis Mór, but it is assumed that the similar landscape features found on Inis Meáin and Inis Oírr will allow for this research to be applicable on those islands as well.

Subject Site - Inis Mór, Aran Islands, County Galway, Republic of Ireland



Appendix A

Nicholas Liquori CZU FES

Coordinate System: IRENET95_Irish_Transverse_Mercator

Chapter 2

Objectives

The goal of this research is to validate the ability to assess habitat quality in semi-natural grasslands of the Aran Islands, Ireland using multi-spectral imagery. If results find that a statistical relationship exists between habitat quality (As defined in the “Grazing Scoring System” of the Agri-Env Scheme Caomhnú Árann) and one or more spatial variables, then it can help identify the most relevant and sufficient variables for predicting priority habitat in pastoral semi-natural grasslands in Northern Europe, and particularly the Aran Islands, Ireland. This will help the plausibility assessment of agri-environmental schemes for these landscapes, assist existing local schemes to expand their project scale and find new participants, and streamline operations and limit input expenses.

Chapter 3

Literature Review

3.1 Caomhnú Árann Operations

The current five-grade scoring system is simpler than the previous system used by AranLIFE, the ten-grade system employed by the successor to BurrenLIFE, The Burren Programme, and other Results Based Agri-Environmental Payments (RBAPS) (Dunford 2020) (McGurn 2020). It is not a continuous scale, but rather ordinally categorizes different fields. Considered criteria when determining a field's score are adequate water supply, controlled scrub or bracken, and most importantly it is the flora biodiversity and existence of key indicator species. Parcels are graded as the units in which they are defined by the Land Parcel Identification System (LPIS) unless further subdivided to correspond with the obvious delineation between grazing scores. If a field is >30% semi-improved/improved land then the two land uses will be scored separately (Caomhnú Árann 2018).

Table 1. List of Indicator Species (Caomhnú Árann)
Source: *www.caomhnuaranneip.ie “Field Scoring System for Assessing Habitat Condition”*

Bird's-foot Trefoil	Spring Gentian	Harebell	Bloody Cranesbill	Pyramidal Orchid	Common Milkwort	Common Spotted Orchid
Early Purple Orchid	Wild Thyme	Eyebright	Devil Bit Scabious	Lady's Bedstraw	Yellow Rattle	Kidney Vetch

Figures 1-2. Examples of grazing score 1 (Caomhnú Árann)
Source: www.caomhnuaranneip.ie "Field Scoring System for Assessing Habitat Condition"

The lowest score of 1 is given to fields that are ungrazed. These are predominately located closer to the shore or along roads and are defined by limestone crags, staging yards, and seldomly tilled land. Most fields receiving a score of 1 will never be eligible for RBAP payments, but where possible, an extensive amount of manual scrub removal will have to take place alongside the reintroduction of grazing and proper provisions of water supply for cattle.

Figures 3-6. Examples of grazing score 2 (Caomhnú Árann)
Source: www.caomhnuaranneip.ie "Field Scoring System for Assessing Habitat Condition"

For scores 2-5, field inspections consider the prevalence of indicator species present in 10 random 1m^2 quadrants and scrub surrounding the fields' perimeter. Fields scored with a 2 characteristically differ from fields scored 3-5 by evidence of soil improvements and summer grazing. In the subject site of this thesis on Inis Mór, these fields are predominantly located along the low laying plain sloping towards the northern shore of the island. They are also in closer proximity to the road and not categorized as winterage (Dunford 2020).

Figures 7-10. Examples of grazing score 3 (Caomhnú Árann)
Source: www.caomhnuaranneip.ie "Field Scoring System for Assessing Habitat Condition"

The largest difference between fields scored 4 and 5 is the percentage of scrub cover. A field scored 5 is allowed to 20% scrub coverage, "as long as it is contained to the edges". Both 4s and 5s should display a high number of indicator species, but there may be trace evidence of slight under grazing on a 4. A field with a score of 3 is grazed, but possesses a lot of issues not present in a 4, but is certainly not improved and as green as a field earning a score of 2. These issues include bare sand caused by overgrazing in Machair habitat or dominance by Molinia. The collection of fields scored with a 3 varies in character much more than the other score groups. See Table 2 for a breakdown of field scores within the subject site of this thesis.

Figure 11. Examples of Molinina Issue in grazing score 3 (Caomhnú Árann)
Source: www.caomhnuaranneip.ie "Field Scoring System for Assessing Habitat Condition"

Figure 12. Examples of Damaged Machair in grazing score 3 (Caomhnú Árann)
Source: *www.caomhnuaranneip.ie "Field Scoring System for Assessing Habitat Condition"*

Figures 13-16. Examples of grazing score 4 (Caomhnú Árann)
Source: *www.caomhnuaranneip.ie "Field Scoring System for Assessing Habitat Condition"*

Figures 17-20. Examples of grazing score 5 (Caomhnú Árann)
Source: www.caomhnuaranneip.ie "Field Scoring System for Assessing Habitat Condition"

Table 2. Summary of fields used in final data frame sorted by grazing score

Grazing Scores in Subject Site					
Grazing Score	1	2	3	4	5
No. of Farms	52	91	47	42	85

Caomhnú Árann is currently undertaking a project of their own to assess the use of UAVs to score habitat quality. This thesis is done independently from Caomhnú Árann, but with overlapping interests and goals. All remote sensing data and on ground data used in the subsequent data analysis was provided to the author by Caomhnú Árann by their own generosity. The author is not aware of Caomhnú Árann's analytical approach and has only been provided data and general information on the organization.

Although the beauty of the islands' landscape comes from its observantly mosaic structure (McGurn 2020), from a classification perspective it is limited. Present on the island is an abundance of limestone, whether in the form of field walls or natural slabs, semi-natural grasslands, scrub, gravel or bituminous concrete drives, and structures. The nature of the conservation scheme makes this research most concerned with vegetated land cover and limestone slabs that exist within grazed areas. These vegetated habitats may be diverse, but this research was not concerned with classifying individual communities or species.

The agricultural practice of winterage is an efficient tradition that has likely been in place since the arrival of neolithic pastoralism(O'Rourke 2005) (Dunford 2020)(McGurn 2020). Its efficiency can only be complemented by a conservation scheme that employs the same efficiency. Caomhnú Árann is a team of three scientists, Scientific and Technical Officer: Amanda Browne, Financial and Administration Officer: Gráinne Ní Chonghaile, and Project Manager: Patrick McGurn, responsible for 125 scattered farm holdings across three separate islands (Caomhnú Árann 2020). It is imperative that they are able to limit expenses and time to secure the financial health of the project and its participants. If on-site field assessment by Caomhnú Árann staff was solely relied upon to provide Grazing Scores, then the project must consider expenses such as overnight accommodation, extensive labor, and nautical transportation fees. To combat these high expenditures of time, money, and resources, Caomhnú Árann is pursuing farmer self-assessment as the next step in RBAS (Browne NRN EIP-AGRI BLOG). Future research is required on the success of a hybrid model where multispectral imagery is used to cross-reference self-assessment by farmers.

3.2 Remote Sensing

GeoAerospace Ltd. was employed by Caomhnú Árann in 2019 to classify fields into grazing scores using imagery from the Sentinel 1 & 2 satellites. The grazing scores given to GeoAerospace Ltd. by Caomhnú Árann were collected in 2016 and thus, satellite imagery from the same year was used in the classification attempts. During preliminary analysis, GeoAerospace Ltd. observed the spectral response variation within each score class to construct an inconclusive classification model and concluded that any machine learning classification attempts would be would not be sufficient if built using the spatial resolution of the Sentinel 2 imagery ($10m^2$) (GeoAerospace 2019).

The recognition of the potential of UAV remote sensing in the field of habitat modeling has been growing in the last 15 years (Lopez 2019). Very often is land which is deemed worthy of conservation in remote peripheral areas, with low funds and resources. UAV remote sensing can save laborious hours of field assessment by small teams in expansive landscapes (Caomhnú Árann 2019). UAVs are being integrated with Real-time Kinematic Positioning technology, which enables surveyors to avoid the preliminary setup of site targets, saving even more time. The exact model used, the DJI Phantom 4 Multispectral RTK, can achieve a horizontal accuracy of 1 cm + 1 ppm, and a vertical accuracy of 2 cm + 1 ppm (Taddia 2019) (Feng 2008). Multispectral data is valuable in assessing grassland habitat for its ability to construct Normalized Difference Vegetation Index (NDVI) (Strong 2017), and the sheer amount of extra data (Komárek 2018). While this thesis did not include the construction of a normalized Digital Surface Model, the results did compare with that of (Komárek 2018) in that classification accuracy was highest when the model was built using different data types (topography and spectral data).

NDVIs are constructed on the fact that vegetation absorbs light within the visible light spectrum for photosynthesis and reflects back infra-red light (Strong 2017). Unlike Strong, this thesis was not attempting to specifically classify vegetation at a community level, but fit a classification system to the Caomhnú Árann Grazing Score System. Strong's research seeks to test the potential of an ENDVI vegetation index to distinguish between different grassland habitats. This index was developed by LDP LLC, Carlstadt, NJ, USA and utilizes the Green and Blue Bands alongside the NIR band (Strong 2017). The equation for this index is

((NIR + Green) – (2 * Blue)) / ((NIR + Green) + (2 * Blue)) (LDP LLC). This Index builds upon the premise that the role of blue light in photosynthesis will “increase the dynamic range of the index values” (Burnside 2020). This approach was not utilized in this research due to the questionable flaws discovered within the values of the blue band. These issues will be described in the Results and Discussions chapter of this thesis. The multispectral data provided for this research did include captured Red-Edge values. There has been suggestion that the characteristics of this spectral band may be beneficial to classifying open landscape vegetation (Strong 2017) (Schuster 2012). While no vegetation index was built using this band, zonal statistics of this band were used in the Random Forest classification all the same as the other 4 bands.

3.3 Random Forests Models

Most Random Forest Classification techniques seek to classify the land cover of individual pixels such as (Svoboda 2022). While this research did experiment with unsupervised classification techniques of the NDVI, the resulting raster was calculated for its zonal statistics per field to be used in as predictors in a Random Forest Model.

This research would like to agree with Zhao in that elevation is the most critical feature for the classification of grasslands (Zhao 2022). However, elevation was not included in any Random Forest Model constructed for this research due to the inability to confirm the continuity of a vertical datum across the three imagery sites. To include some bit of topographic data into the classification attempt, slope was calculated as a method of normalizing the surface elevations.

The number of trees (ntree) parameter within the random forest algorithm was set to 1000 across all models. Xu scaled this parameter to the “relatively high value of 500” until the smallest OOB error was achieved (Xu 2019). This method was also applied with this research. The mtry parameter was set to default (square root of number of predictors) as recommended by the ranger packet documentation (Wright 2021).

Chapter 4

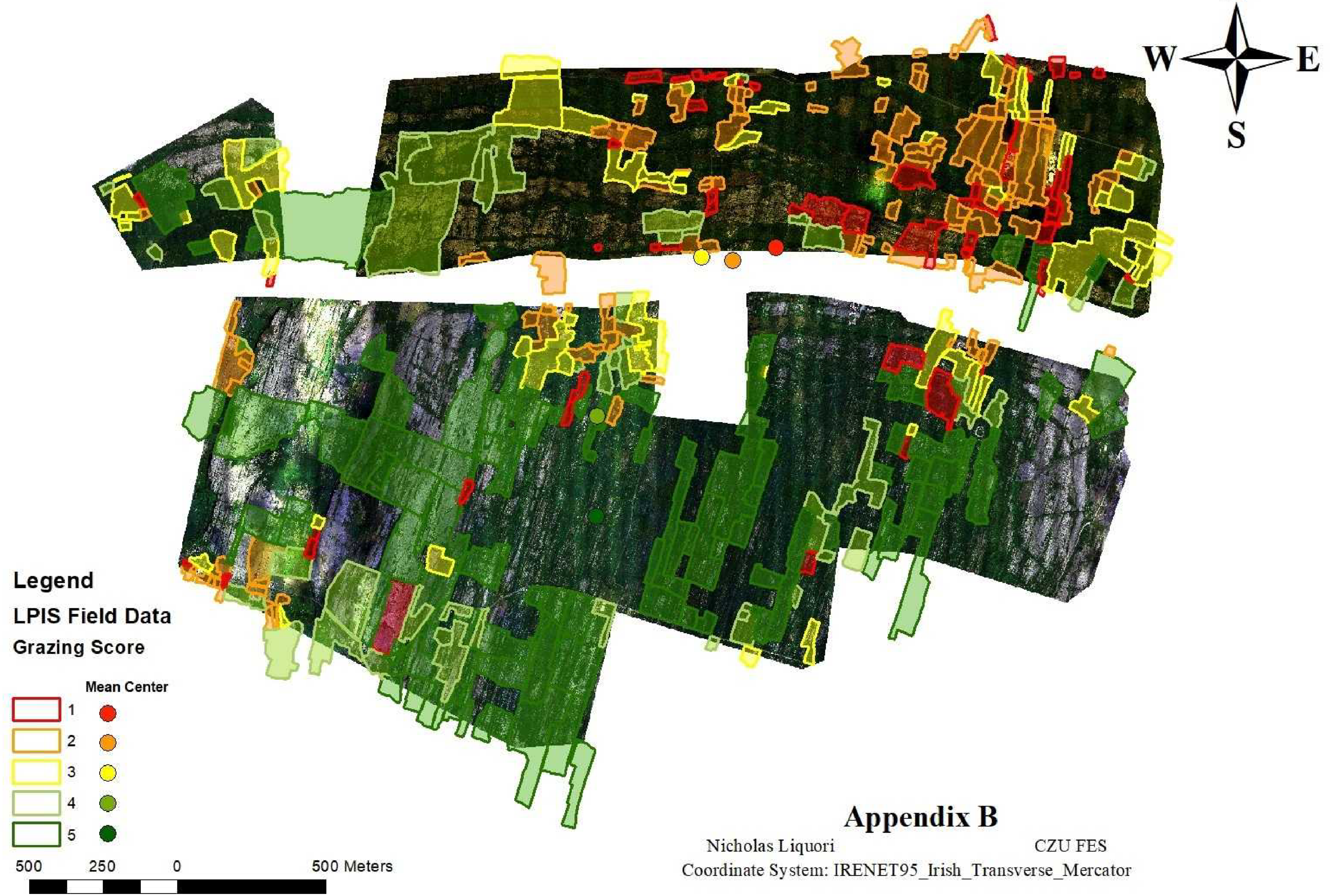
Characteristics of the Subject Site

The Aran Islands are a chain of three islands located off the west coast of Ireland. Across Galway Bay to the NE is County Galway, and directly east is located The Burren in County Clare. The three islands named Inis Oírr, Inis Meáin, and our subject site Inis Mór, are defined by rare Karst limestone landscapes and orchid rich calcareous semi-natural grasslands (Colgan 1893). The islands also host the southerly limit of the geographic range of the rare machair habitat in Europe (Bassett 1985). These habitat types have developed a dependency on agricultural management. Threatened by agricultural polarization (abandonment ---- intensification), conservationists in The Burren and The Aran Islands have opted for a subsidy scheme that provides farmers with payment for corrective action and habitat maintenance. The efforts have changed faces through different EU subsidy programs, but their purposes have always remained the same.

The limestone geology was formed during the Carboniferous, on the bottom of the tropical sea (Self 1998). The islands share the same geological history as the Burren, which Drew titles “the finest example of a karstic landscape in Ireland” (Drew 1997). Despite its rugged exterior, the Islands are very much so an agrarian landscape. Domesticated grazing animals have been present since expansive clearing occurred in the Mesolithic-Neolithic transition (Molloy 2004). Since this time, the prospering grasslands and flora species developed a strong dependency on constant grazing levels. Many of these rarities and beauties present in the cultural landscape have made the islands a target of conservationists and tourists.

Figure 21. SE corner of Inis Mór, located off subject site

Grazing Scores of Subject Site Farms



Appendix B

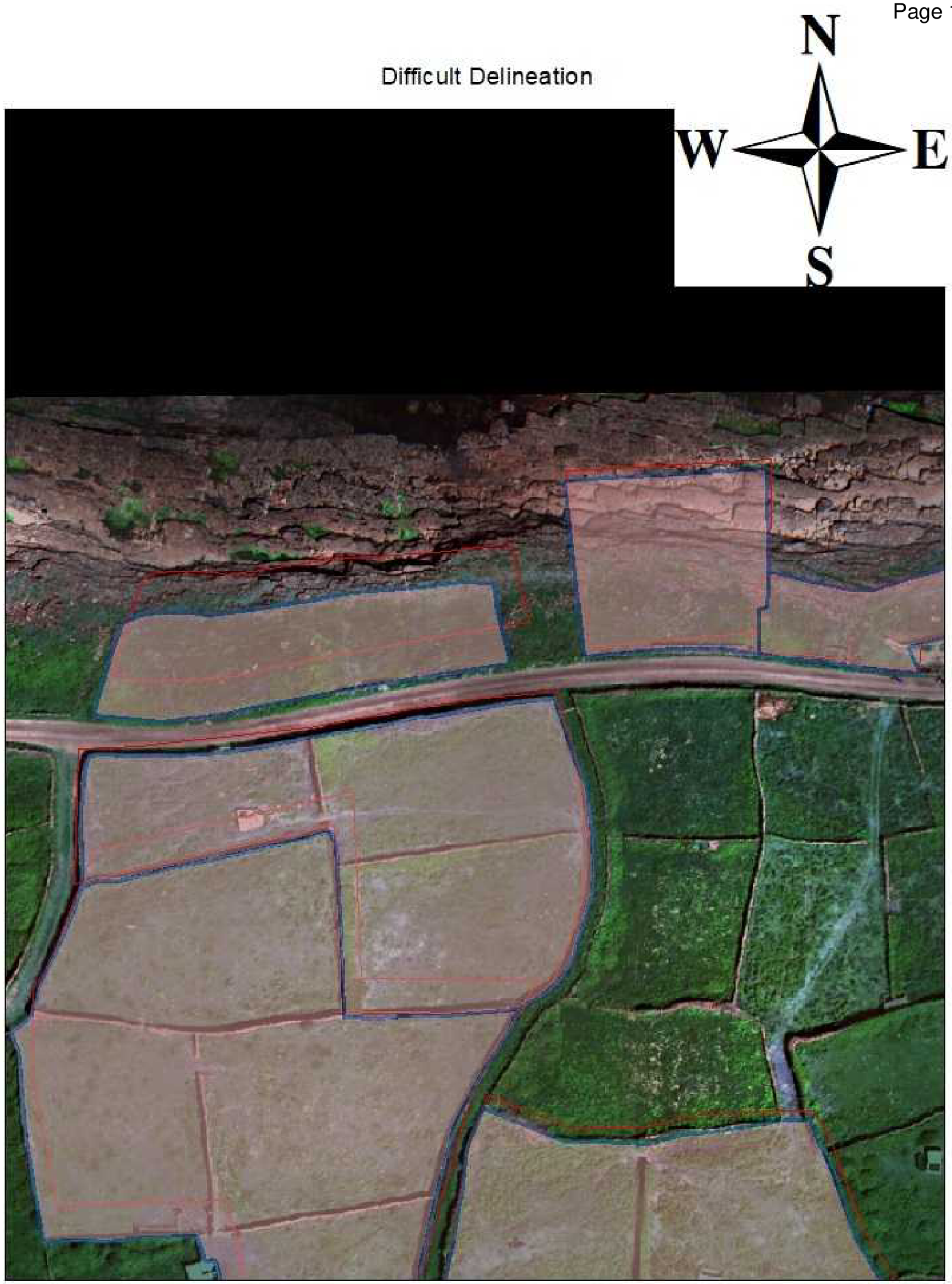
Nicholas Liquori CZU FES

Coordinate System: IRENET95_Irish_Transverse_Mercator

Edited LPIS Field Data

Easy Delineation

Difficult Delineation



20 10 0 20 Meters

Appendix C

Nicholas Liquori CZU FES

Coordinate System: IRENET95_Irish_Transverse_Mercator

Legend

Edited LPIS Field Data

Original Shapefile

Chapter 5

Methodology

5.1 Data

Caomhnú Árann employed the services of GeoAerospace Ltd. to conduct a remote survey of three adjacent sites on Inis Mór. The sites were flown on June 28th & 29th, 2021 with a DJI Phantom 4 Multispectral UAV and output rasters were ortho rectified by a built in RTK GPS. The DJI Phantom 4 Multispectral UAV utilizes photogrammetry to map the horizontal and vertical coordinates, and produce the output products (point cloud, DEM, and 5 band Orthomosaic). The band names and their spectral scale acquired by this drone survey are as follows: Blue (B): 450 nm ± 16 nm; Green (G): 560 nm ± 16 nm; Red (R): 650 nm ± 16 nm; Red edge (RE): 730 nm ± 16 nm; Near-infrared (NIR): 840 nm ± 26 nm. The drone was flown at 120m above ground level for each flight. The vertical datum references the OSGM 1.5 Model and elevation values are in ellipsoidal height (Foyle 2022).

The three sites known colloquially as Mór 1,2, & 3 were flown at separate times across June 28th and 29th, and do not geographically overlap at all. The sites are numbered in counterclockwise order starting from the small NW site. The order in which the editing of the shapefile was conducted corresponding to their respective sizes: Mór 2, Mór 3, and finally Mór 1. Editing procedures were kept constant across the entirety of the data. Appendix B shows the geographic limits of all provided raster and vector data.

Preparation for data analysis began by creating analytical rasters and editing the provided shapefile. GeoAerospace provided the orthorectified 5 band imagery and a DEM, and Caomhnú Árann provided the field data shapefile hereon titled "LPIS Field Data". The LPIS Field Data shapefile was edited so that polygons were topologically flush with neighbors where true and were most accurately representing the field walls visualized in the remote sensing imagery.

5.2 Remote Sensing Data Preparation

Various tools were needed to acquire the desired input predictors and manage the data to be processable. Every individual band was extracted as a lyr. file and then exported as geotiffs. An issue was noticed early that the areas surrounding the extent of raster data which should be "noData" were in fact valuing as 0. This was handled by using a Con expression in the raster calculator with all five individual bands as input rasters. The resulting raster was one which valued all pixels possessing a value of 0 in all five individual bands from the multispectral imagery as 1. Upon this processing, it was discovered that .002 % of the data area was in fact containing no data from the drone flight. This is a result of a lack of ambient light penetrating shadowed areas on the drone flight path, and thus the drone camera detected no light coming back. The uncovered area was considered insignificant due to small coverage.

The output raster from the Con expression was then used as the conditional raster in a SetNull expression with each of the five multispectral bands being used as constants. This resulted in five individual single band rasters representing the five spectral ranges, valuing as NoData instead of 0 surrounding the data extents. The five individual bands were combined back together using the "Image Analysis-Composite Bands" feature in ArcMap. The RGB imagery now valued all pixels outside the true data extent as NoData. This data management was necessary in order to accurately calculate an NDVI. The DEM provided by GeoAerospace was already satisfactory in this regard. A raster representing slope in degree was created from the DEM.

The NDVI was created with the following equation (Red-NIR/Red+NIR) (Sun 2014). The Red Band was the third band provided by GeoAerospace with a spectral range of 650 nm ± 16 nm, and the Near-Infrared Band was the fifth band with a resolution of 840 nm ± 26 nm. AN NDVI raster was provided for this thesis research by GeoAerospace but was not used in further analysis. It was however compared to the NDVI created as a part of this research. Differences in the two data sets were negligible to none.

To capture the distribution patterns of vegetated characteristics inherent with each of the grazing scores, further analytical rasters derived from the NDVI were created. The NDVI was used as the input raster in 3x3 rectangular focal statistics. This operation was run for the following statistical types resulting in four new rasters: Mean, Standard Deviation, Percentile, and Range. The NDVI was also used as the input raster for operating a smoothing low pass filter and edge-enhancing high pass filter.

5.3 LPIS Data Preparation

The LPIS is the cadaster database used by DAFM to manage payments to their area-based payment schemes. The current system is 25 years old and the newly introduced system created in 2019 has not been mainstreamed in the Aran Islands' County Galway (DAFM) (LPIS 2019). This may explain to the crudeness of the geographic accuracy and topological inconsistency in the provided shapefile. Appendix C shows the manual transformation of the LPIS Field Data polygons done for this thesis. The goal for this editing procedure was that the limits of each polygon were placed in the center of the field wall they corresponded to, that adjacent fields were flush to each other where they were supposed to be, and no two polygons would overlap.

Decisions about the "true" location of polygons vertices were made using best judgment, as many field walls are obvious demarcations of parcel limits. The easiest fields to delineate were those whose borders were demarcated by well-managed field walls or by roads, as seen in Appendix C. Some characteristics of difficult delineations include fields whose rear boundary was defined by severe slope and large irregular limestone slabs, or residential lots that staged agricultural equipment around its perimeter. An example of some unobvious field boundaries can be seen in Appendix C. Extreme conservatism extra care was used in judging the perimeter of these parcels by retaining as much of the original geometry as possible. To make accurate edits, but not get overwhelmed by the intricacies of field wall pivots, the data was usually scaled between 150 – 250:1m. Vertices were always placed on the best approximation of the center of field walls. The polygons were displayed with a degree of transparency and proper symbology to ease in the visualization of the rasters they were to be corresponding to. In ArcMap, the end vertices of adjacent fields would be snapped together, and then the edge that they were supposed to share was snapped together using topology tools. The resulting shared edge would be edited together, again using tools from the same topology toolbox. Keyboard shortcuts were created in ArcMap to speed up the editing workflow. The topology tool was used upon completion of each site and then upon completion of all sites to scan for any polygons that were not flush as they were meant to be. Where polygons extended past the limits of the remote sensing data, vertices were snapped at the intersection of their “true” location on field walls and the limit of the raster. No polygon edge was edited outside of the geographic limit of raster information. It should be noted that the metric for field area was recalculated after the editing procedure of the LPIS Field Data Shapefile. This shapefile has been shared with Caomhnú Árann to aid in any future spatial analysis done on Inis Mór.

5.4 Predictors

The final data frame used for the “Large Model” and the Principal Component Analysis included 317 observations of 48 variables. The variables chosen source from the Individual Bands, NDVI, NDVI derived rasters, DEM, and the LPIS Field Data. Mean, Range, and SD Zonal Statistics were calculated amongst each field polygon for the NDVI and its derivative rasters, Mean, Median, Range, and SD for the individual bands, Mean, Range, SD, Min, and Max for the Slope raster, and Area for each polygon was also calculated. Caomhnú Árann provided a shapefile containing 369 observations. 3 of these observations were removed because they were entirely outside of the extent of remote sensing data, and 49 other observations were removed because they met an arbitrary threshold of percent coverage of remote sensing data chosen for this thesis. Fields whose area was not at least 80% covered by the remote sensing data were removed from the subsequent statistical analysis.

Table 3. List of predictors and their source data

MS Bands	**Blue**	**Green**	**Red**	**Red Edge**	**Near Infrared**
	Mean	Mean	Mean	Mean	Mean
	Standard Deviation	Standard Deviation	Standard Deviation	Standard Deviation	Standard Deviation
	Range	Range	Range	Range	Range
	Median	Median	Median	Median	Median
NDVI	**Original NDVI**	**Focal - Mean**	**Focal - SD**	**Focal - Percentile**	**Focal - Range**
	Mean	Mean	Mean	Mean	Mean
	Standard Deviation	Standard Deviation	Standard Deviation	Standard Deviation	Standard Deviation
	Range	Range	Range	Range	Range
	High Pass Filter	**Low Pass Filter**	**Surface Slope**	**LPIS Field Data**	
	Mean	Mean	Mean	Area	
	Standard Deviation	Standard Deviation	Standard Deviation		
	Range	Range	Range		
			Min		
			Max		

The LPIS Field Data also included a code representing farm ownership. There were 35 unique farm codes representing farm ownership. Figure 37 shows the relationship between field size and score but averaged per farmer. This visualization shows that there is not much variability in the case from farmer to farmer and that the performance of an individual farmer will not affect the classification model. For this reason, Farm Code was excluded from the final data frame. Farms are scattered as can be seen in Appendix D. It shows the dispersion of the 10 farmers with the largest number of separate holdings. Individual farmers have a diverse landscape character across their fields and typically perform across the scoring system spectrum.

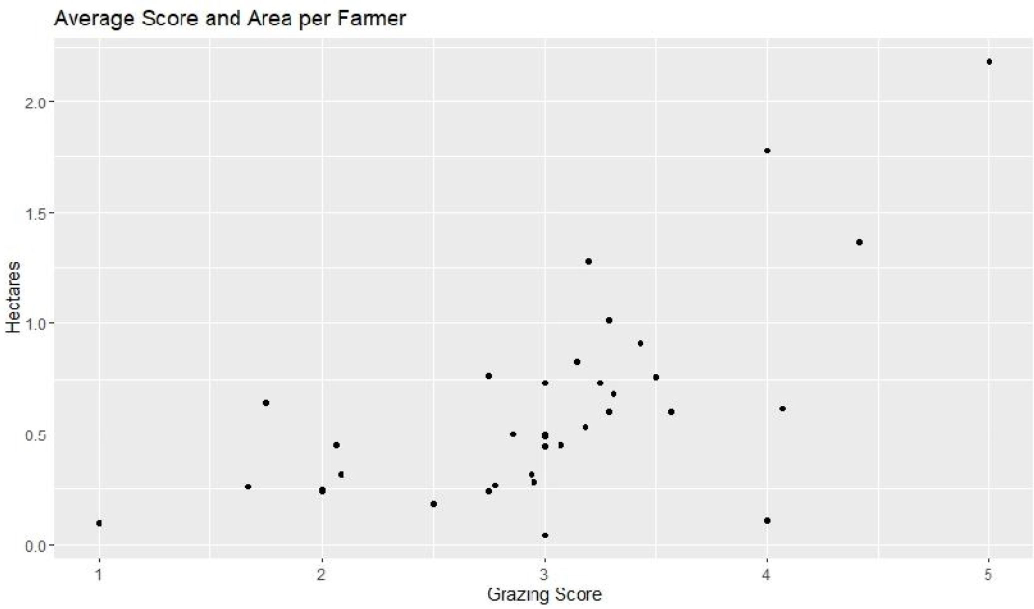


Figure 22. Relationship between area and grazing score averaged per farmer

5.5 Exploratory Analysis

Exploratory analysis was done by making Box Plots of the area of LPIS farm units, and the zonal mean and standard deviation for each of the input rasters. These variables were taken from the data frame shown in table 3. The data was summarized using the field scores as a case field. The input rasters were created from the multispectral imagery and the zonal statistic variables were created using ArcMap 10.8.1 for Desktop. The accompanying boxplots and all data analysis in this thesis were done in R studio using the programming language R. Variable width displays the width of each candle as they represent the percentage of observations and the color of each candle matches the symbology used for each grazing group across the entirety of this thesis.

5.5a Correlation Matrix

Due to the nature of the input rasters, a significant level of collinearity was expected. A correlation matrix was constructed to observe this in preparation for model refinement. The actual variables which constitute the data frame are zonal statistics calculated in GIS for each field polygon. The correlation matrix was built using the corrplot library in R.

5.5b Principal Component Analysis (PCA)

Multi-collinearity of input variables can make the interpretability of variable importance in classification models quite difficult. An important goal of this thesis is not only to attempt classification but tell Caomhnú Árann what inputs are relevant in future remote sensing surveys. A Principal Component Analysis helped achieve this goal. All input data shown in table 3 was also used in a Principal Component Analysis.

5.6 Iso-Cluster Unsupervised Classification

Because Random Forest is a supervised classification technique, this thesis hoped to diversify its contribution to Caomhnú Árann's efforts by also including an unsupervised classification technique. The unsupervised classification method used was Iso-Cluster Unsupervised Classification. This method uses an isodata clustering algorithm to determine the characteristics of the natural groupings of cells (Lemenkova 2021). This operation would be used to create a 3 class raster, 5 class raster, and 10 class raster. The 3 class raster was classified using only the NDVI as an input raster, while the 5-class and 10-class rasters were constructed using the NDVI and Slope as input rasters. Default values for this unsupervised classification tool in ArcMap were used. These include the number of iterations of the clustering process to run (20), the minimum number of cells in a valid class (20), the interval to be used for sampling (10). No signature file was created in ArcMap because the supervised classification was to be done in R using the Random Forest Model.

5.7 Random Forest Classification

The decision to build a random forest model was made for its ability to classify ordinal data and ease of construction and understanding. Random Forest is a common method for supervised classification. These models are a defined number of decision trees that randomly generate vectors from the input data frame for each tree. The most popular result for each observation across every tree elects its predicted class.

5.7a PCA Random Forest Model

Principal component analyses plot input variables in as many dimensions as there are input variables, and project the data onto new components. The result will provide you with as many components as you provided input variables, but with a satisfactory coverage of explained variance in fewer variables. The new components will reflect the same dataset, lessen collinearity present in the original dataset, and decrease the number of variables in any further modeling. In the case of this thesis, the 47 variables which built the large random forest model were transformed, resulting in 47 components. The first ten components were selected and bound to the original observations, and used as predictors in

All parameter settings remained constant across all six Random Forest Models, but mtry, which stands for the number of predictors from which a best split is chosen in each decision. The default value for mtry in the ranger package is the (rounded down) square root of the number variables in the model. This default value was used in each model. For this model, with 10 input variables in total, the mtry value is 3. The minimum size for terminal nodes in this and all Random Forest Models in this thesis research is 1. This is recommended for classification models in the documentation published for the R package Ranger (Wright 2021).

5.7b Large Model

The largest Random Forest model built in the effort of this research, hereon titled "Large Model", was constructed with every input variable used in the Principal Component Analysis. . For this model, with 47 input variables in total, the mtry value is 6.

5.7c Small Model

The variables chosen for the more selective smaller model were derived from the original data set. There were no zonal statistics chosen from the same input raster and all variables chosen possessed a level of variable importance in the "Large Model" greater than the median. The input variables can be seen below in table 4.

Table 4. List of predictors and their source data for smaller selective model

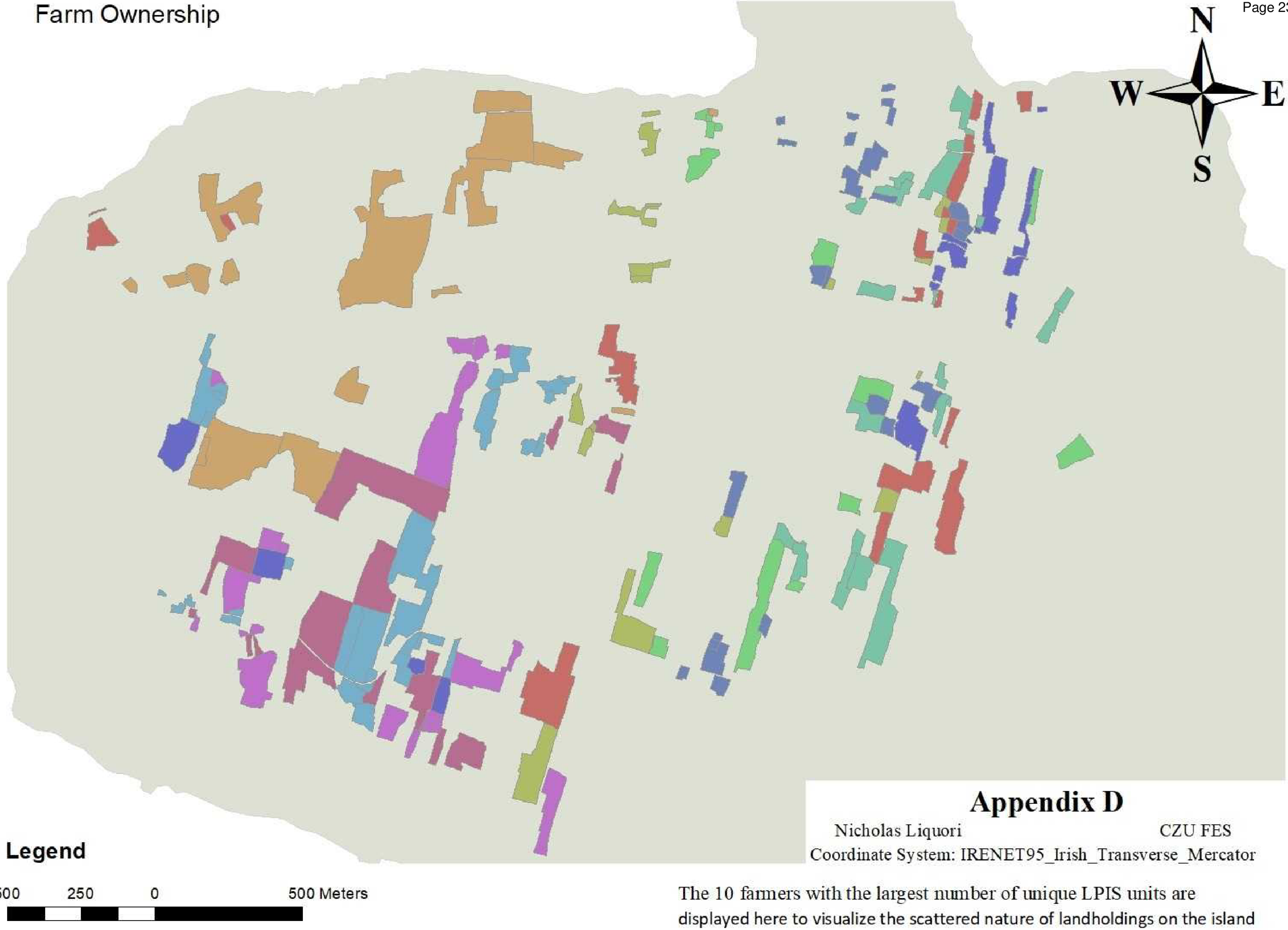
MS Bands	**Green** Median	**Red Edge** Median	**Near Infrared** Mean		
NDVI	**Original NDVI** Standard Deviation	**Focal - Mean** Standard Deviation	**Focal - SD** Mean	**Focal - Percentile** Standard Deviation	**Focal - Range** Mean
	High Pass Filter Standard Deviation	**Low Pass Filter** Standard Deviation	**Surface Slope** Min	**LPIS Field Data** Area	

The purpose of this smaller model was to evaluate if the variables with the greatest variable importance in the "Large Model" could construct a classification model with better prediction accuracy. The goal of this research is to give Caomhnú Árann the most information possible. Analyzing results from random forest models with different inputs will achieve this goal. For this model, with 12 input variables in total, the mtry value is 3.

5.7d Percent Area of Unsupervised Classes – Random Forest Models

The percent area of each of each class in the resulting rasters was calculated per LPIS field unit. The method for doing so included rasterizing the LPIS Field Data and setting the pixel size to match that of the classification raster, and then using the "Tabulate Area" function in GIS. To assure proper calculation of area %, the sum of the area of the classes present in each polygon was used as the denominator, in contrast, to simply using the total area of the polygon itself. The resulting variables which were used in Random Forest Classification were % Class 1, % Class 2, … % Class 10 of each LPIS field unit. For these models, with 1, 5, and 10 input variables, the mtry values were 1, 2, and 3.

Farm Ownership



Appendix D

Nicholas Liquori CZU FES

Coordinate System: IRENET95_Irish_Transverse_Mercator

The 10 farmers with the largest number of unique LPIS units are displayed here to visualize the scattered nature of landholdings on the island

Chapter 6

Results

6.1 Exploratory Analysis

As hypothesized, NDVI was significant to the grazing score of individual fields. Fields scored 5 possessed a higher standard deviation but a lower mean of NDVI values across a single field. This thesis also paid attention to the diversity of green candles across the five grazing scores but could not identify any sign that green values are highest in mixed grasslands (He 2006).

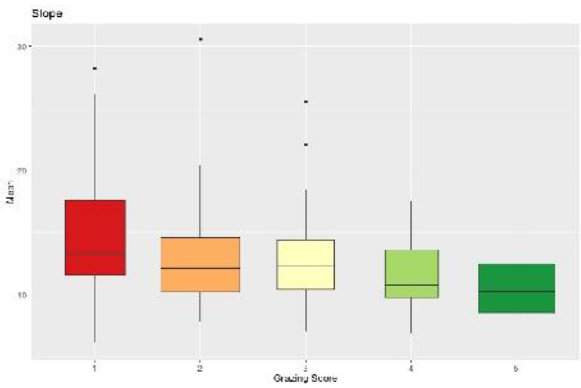


Figure 23. Slope – Zonal Mean

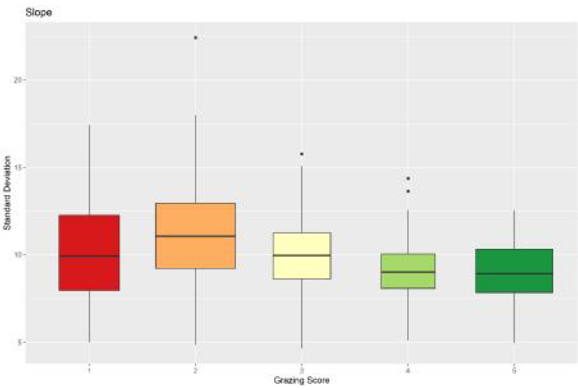


Figure 24. Slope – Zonal SD

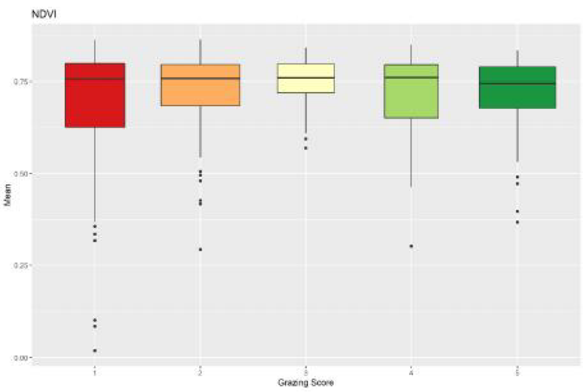


Figure 25. NDVI – Zonal Mean

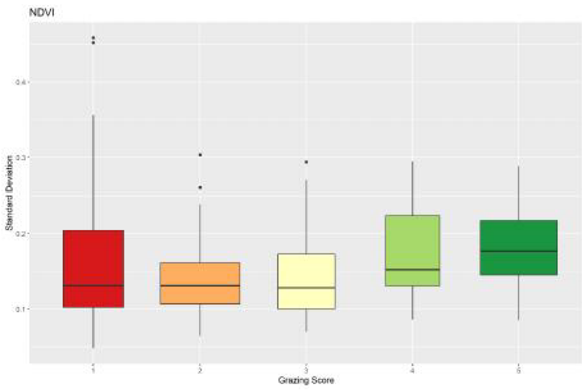


Figure 26. NDVI – Zonal SD

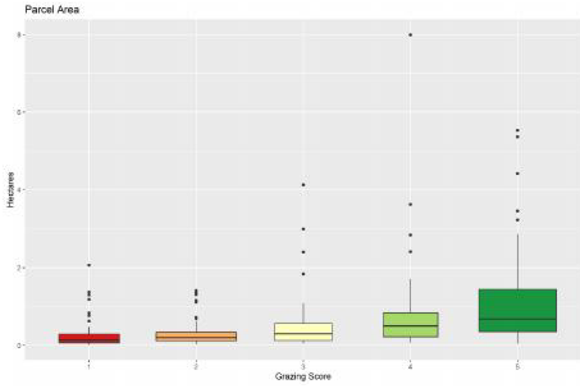


Figure 27. Field Area

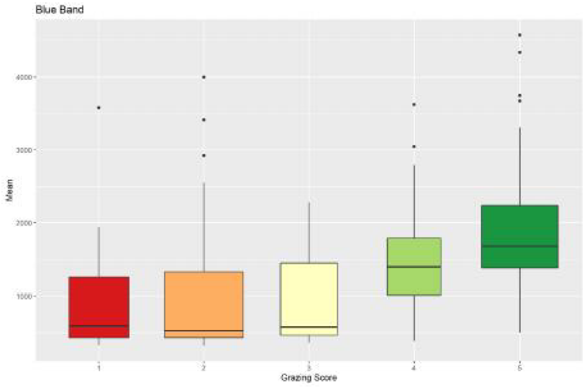


Figure 28. Blue – Zonal Mean

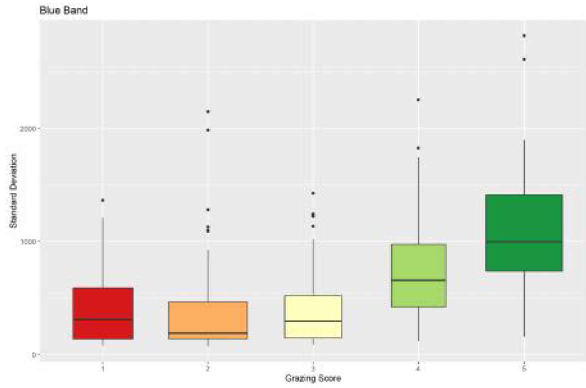


Figure 29. Blue – Zonal SD

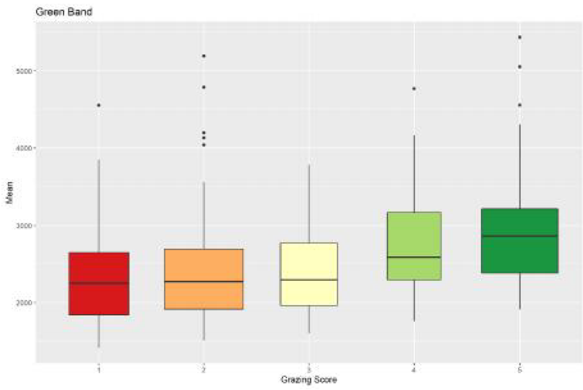


Figure 30. Green – Zonal Mean

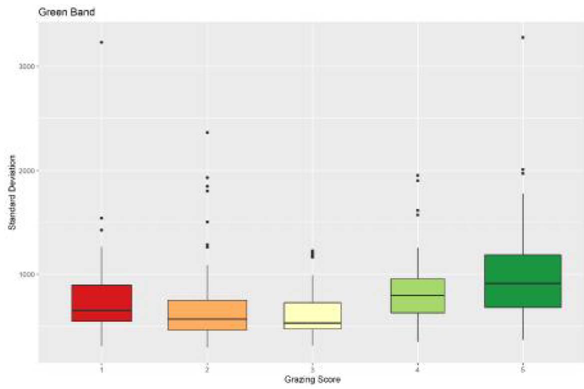


Figure 31. Green – Zonal SD

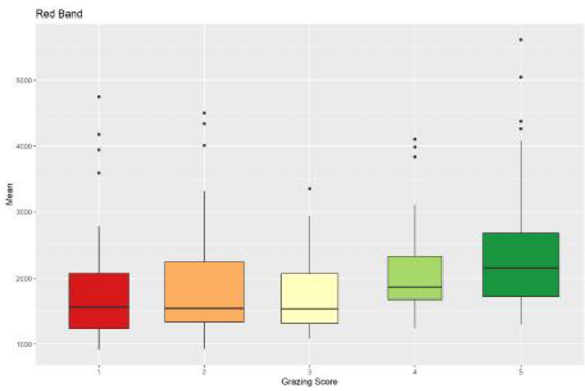


Figure 32. Red – Zonal Mean

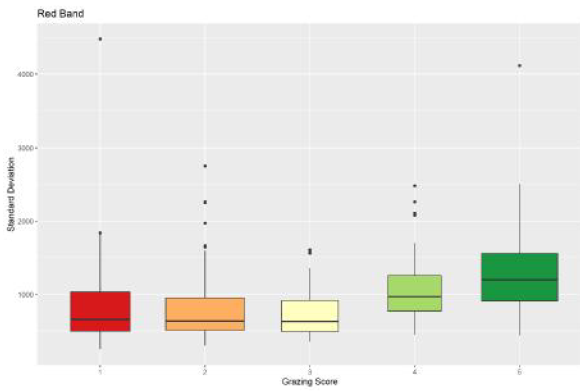


Figure 33. Red – Zonal SD

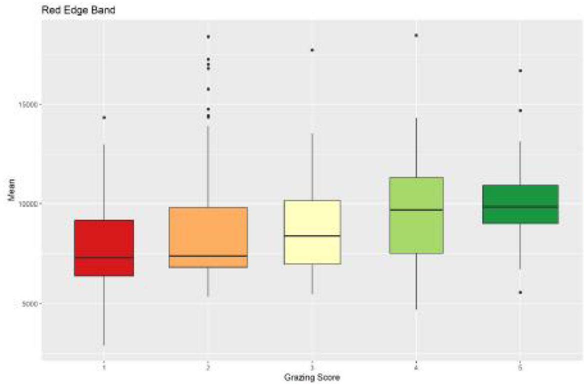


Figure 34. Red Edge - Mean

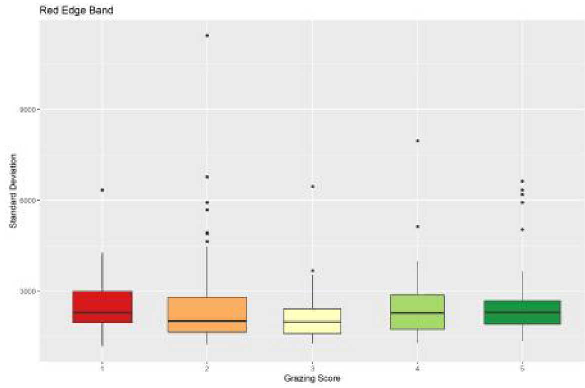


Figure 35. Red Edge - SD

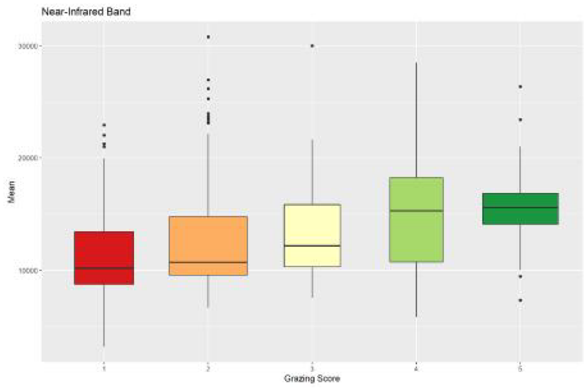


Figure 36. Near-Infrared – Zonal Mean

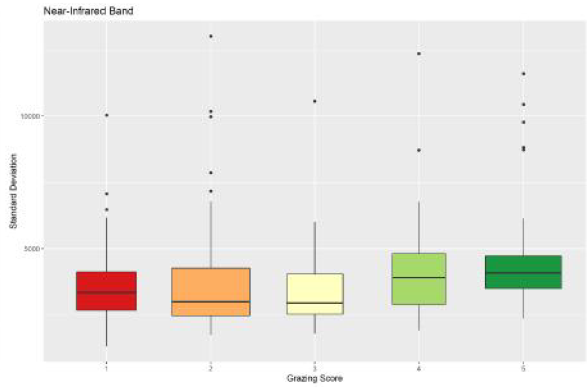


Figure 37. Near-Infrared – Zonal SD

The group of fields scored 5 has the highest mean Standard Deviation of NDVI values as well as the tightest range. While the group mean of mean NDVI values shows no clear pattern, the Standard Deviation and Range of these values for the group of fields scored 3 does match with the hypothesis of this thesis. The mean values for slope across the 5 groups also matches the geographic distribution of field scores across the subject site. There is a general directional trend for the mean geographic center of each score group as shown in Appendix B. Due to the concentration of samples in the NE corner of the subject site, there is a general sway of the mean centers towards that corner, but most promising is the obvious sway of mean centers from 5-1 moving in a northerly direction towards the coast. The winterage, which is mostly made up of 5s and 4s, is located on the high flat plateau in the central belly of the island. Most of the lower-scored fields made up of improved land are located on the northern sloping plain up to the coast. These results match the general characteristics of these score groups.

Figures 38 and 39 display the box plot for the zonal statistics of the raster output from the Mean Focal Statistics for the NDVI. These results are very close to those of the box plots for the original NDVI. One can expect these two data sets to be highly correlated. The remainder of the exploratory box plots of predictors can be found in the appendix.

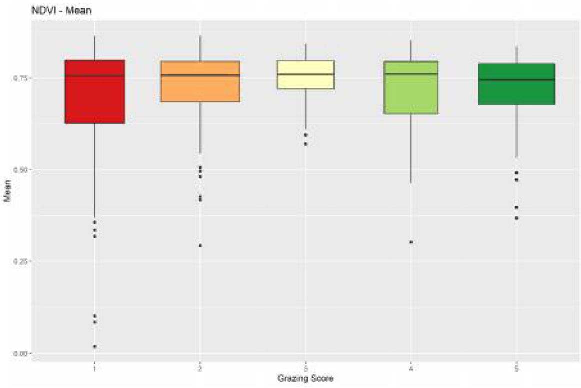


Figure 38. Focal Mean NDVI – Zonal Mean

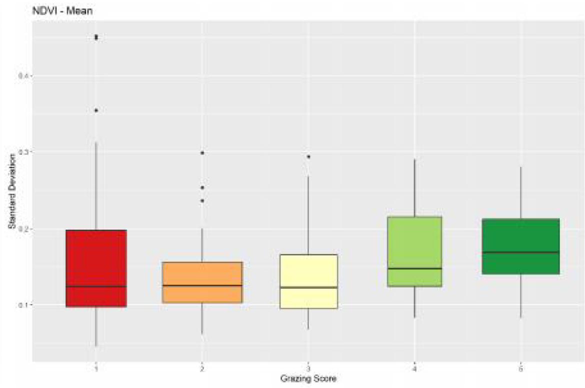


Figure 39. Focal Mean NDVI - Zonal SD

6.2 Correlation Matrix

The results from the Correlation Matrix proved the high collinearity of input variables. Zonal means for NDVI derived variables had a strong positive correlation with other zonal means from the same derivation and a strong negative correlation with zonal Standard Deviation from NDVI sourced variables. Slope and Area were uniquely sourced and thus showed the relatively small correlation coefficients expected of them. Amongst variables sourced from dissimilar rasters, the highest level of collinearity was between variables derived from the Red band and the zonal Mean and Standard Deviation of NDVI sourced rasters. This does not come as a surprise as the NDVI is calculated from the Red Band. See Figure 39 for the Correlation Matrix. These results came into consideration when limiting the number of inputs into a smaller random forest model.

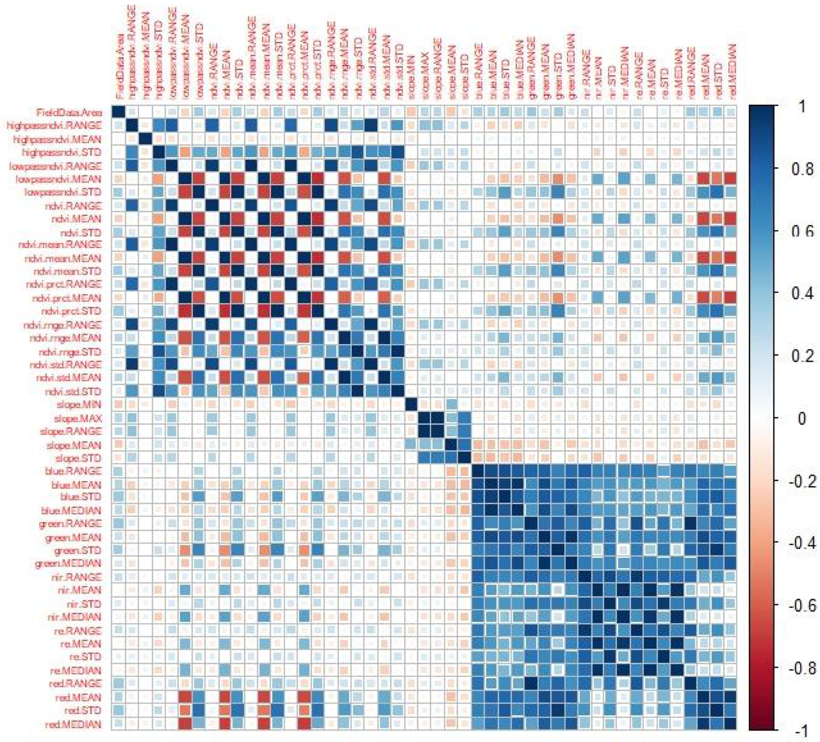


Figure 40. Correlation Matrix of all input variables

6.3 Principal Component Analysis

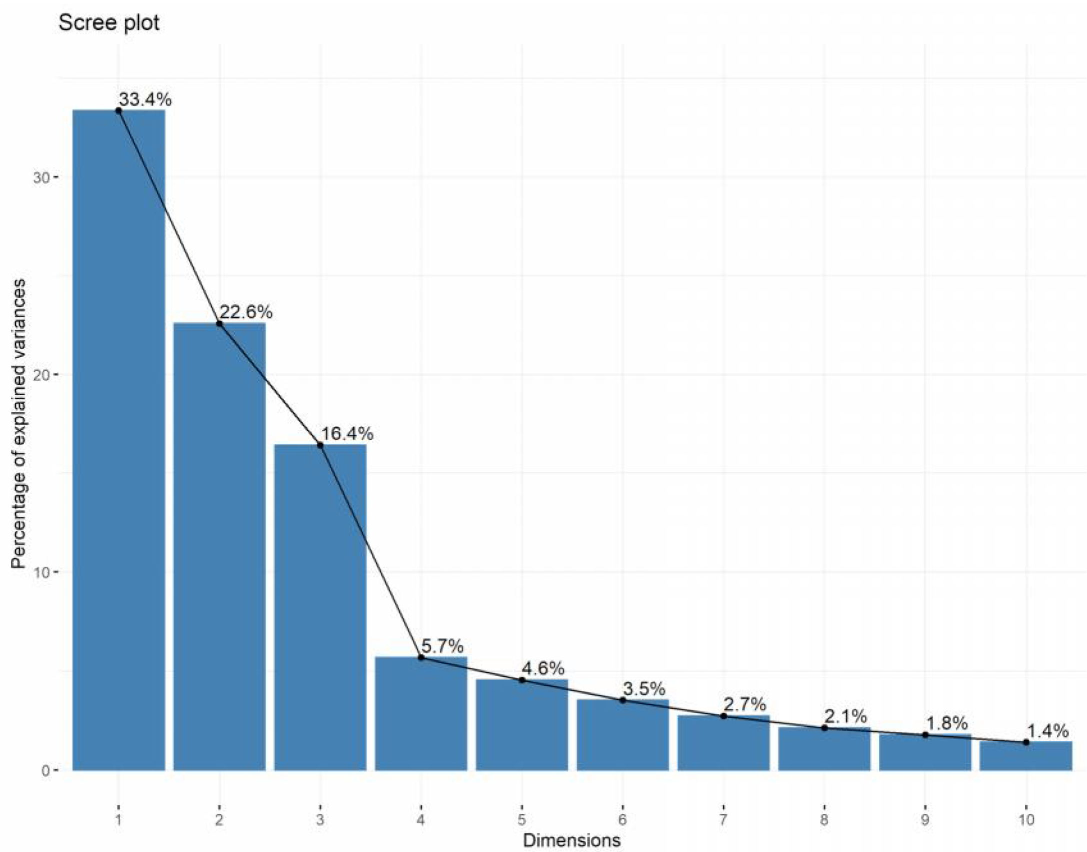


Figure 41. Scree plot of explained variance per component

The resulting scree plot, which shows the percentage of explained variance in each component, can be seen in figure 41. This displays the percentage of explained variance projected onto each of the first ten components. The cumulative proportion of variance present in the first 10 components is 94.151 %. This thesis was always going to be satisfied with a stopping point of 90% cumulative proportion of variance and the 10^{th} component also satisfies the Guttman-Kaiser criterion which is to analyze components that have an eigenvalue greater than 1 (Jackson 1993) (Jackson 2003). The overall goal of the principal component analysis was to remove collinearity in the data, discover which variables were most important, and provide an alternative to a large random forest model built by an overwhelming data frame. The ten principles were bound to the source data and used as predictors in a Random Forest Classification model.

The input variables and their contribution percentage of explained variance within the first two component dimensions can be seen in FIGURES XX and XX.

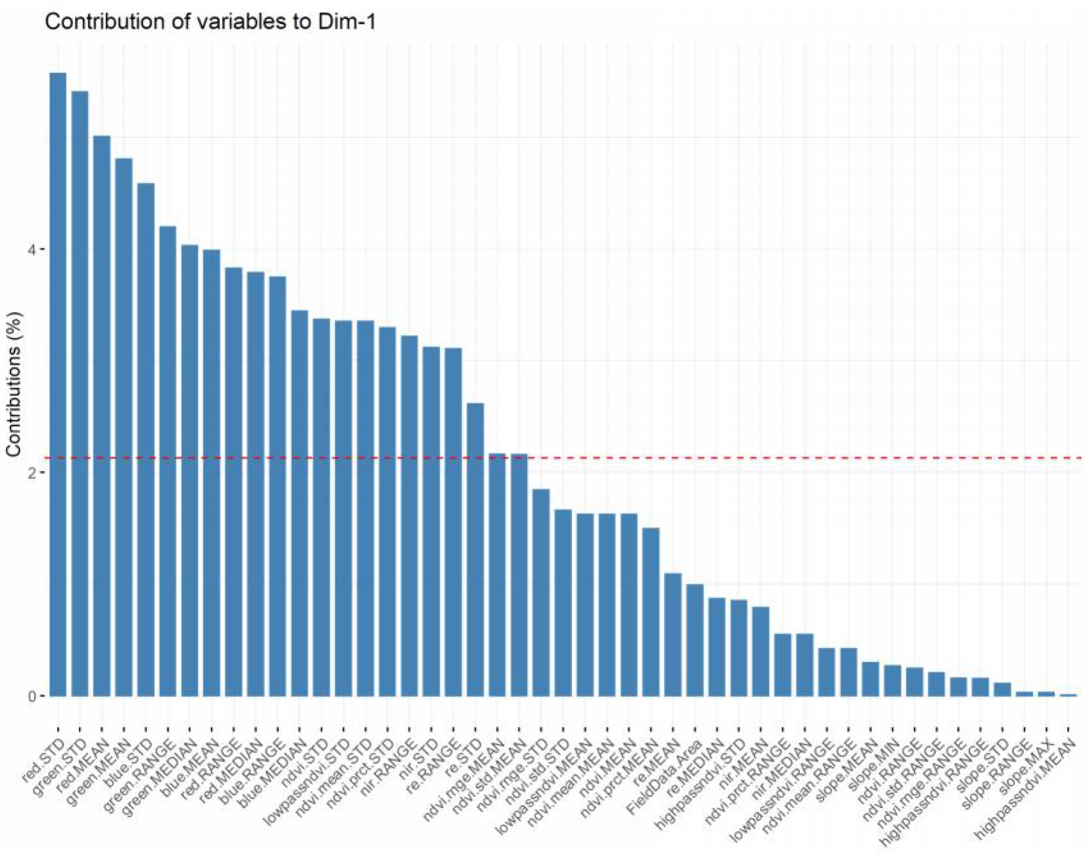


Figure 42. Contribution of variables to dimension 1

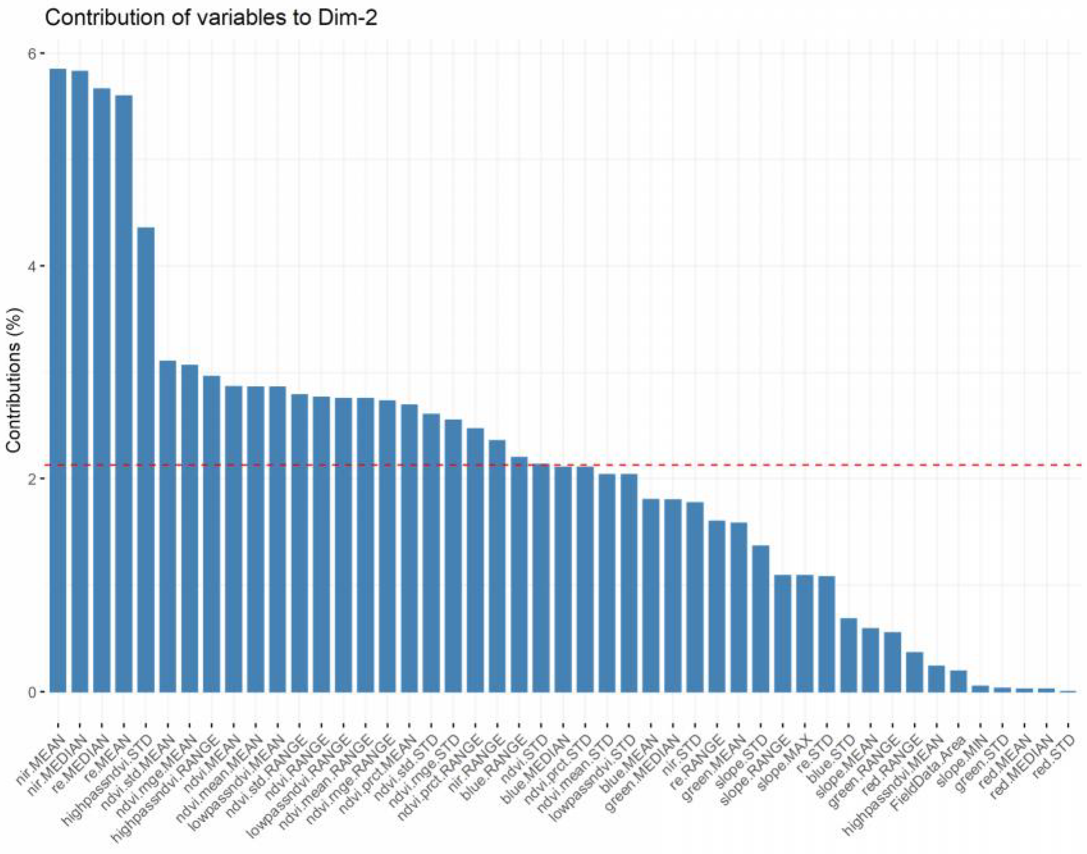


Figure 43. Contribution of variables to dimension 2

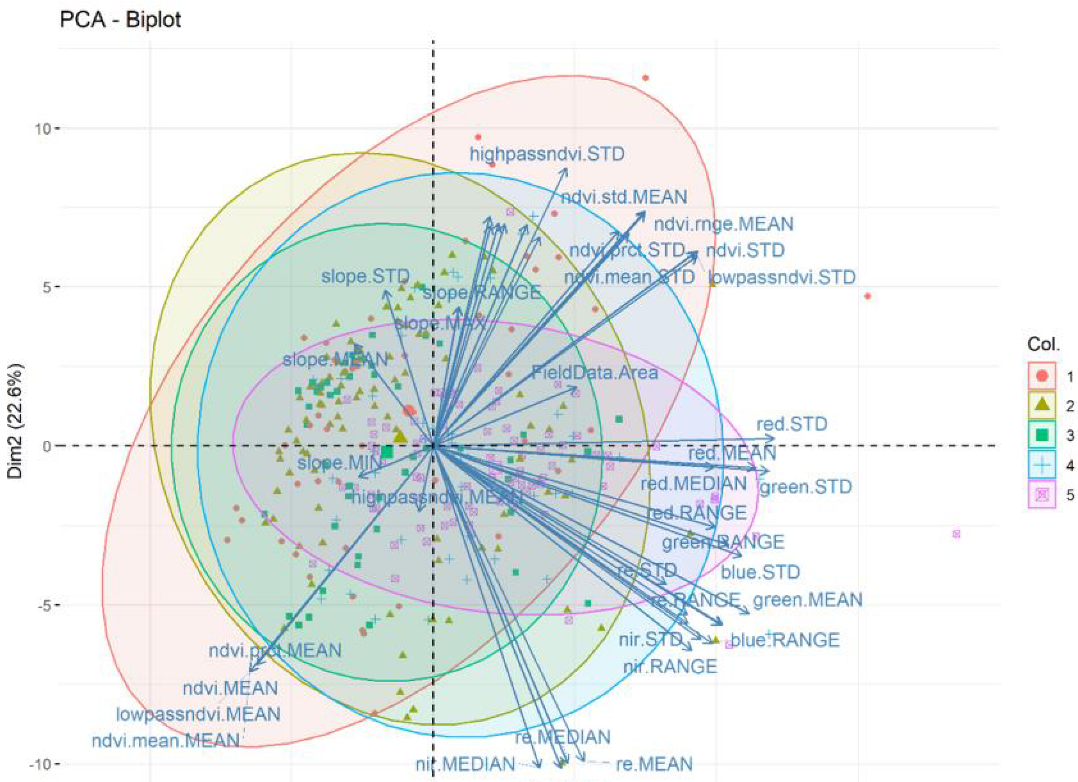


Figure 44. Bi-Plot of principal component scores and variable loadings

The Biplot (figure 45) output from the PCA tells a lot about the input variables' contribution and variability exhibited in the first two components. Additionally, there is evidence pointing to which variables are most relevant to specific grazing scores. The variability of variables derived from the individual bands of the multispectral imagery appears to be represented very well in the first two components, and their contribution to the first component specifically is visible by its easterly orientation on the x-axis. In general, the variability from remote sensing imagery (Individual Bands and the derived NDVI) is expressed well in the first two components in contrast to field area and variables derived from the terrain slope (See FIGURE XXXX). Variables derived from slope do appear to have a strong negative loading on component 4 as seen in FIGURE XXXX. Clustering amongst grazing scores is not obvious in the biplot of the first two components, but perhaps the tight ellipses of the score 5 observations can hint at success in the ability to predict these farms in the random forest model using the first two components.

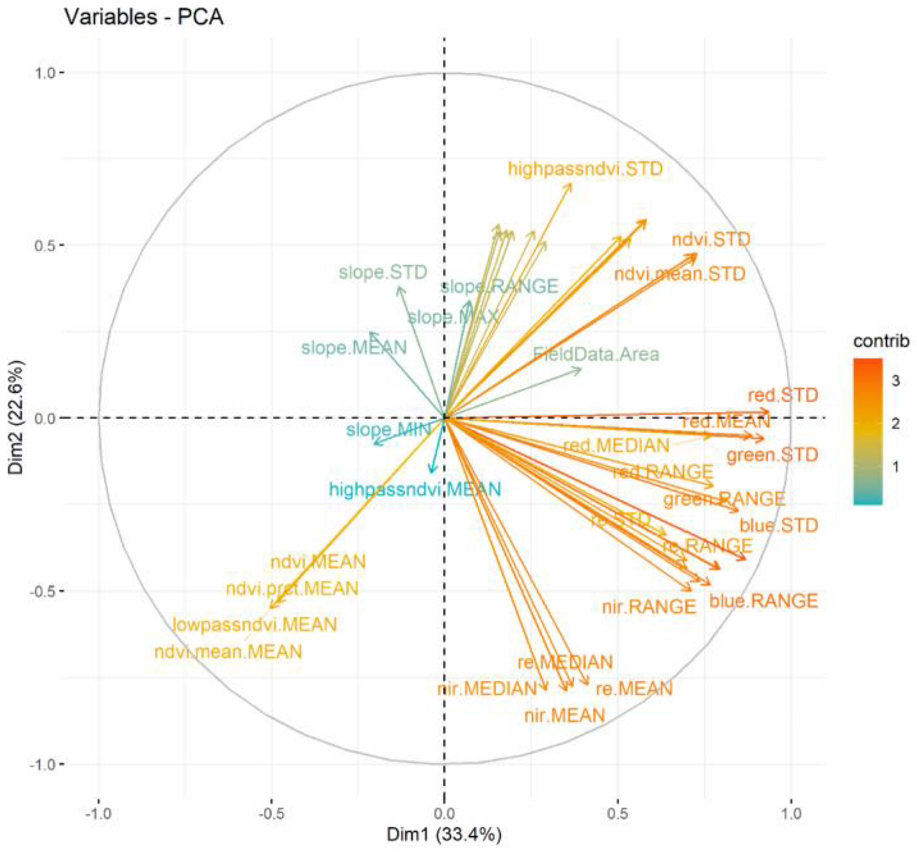


Figure 45. Loading plot of first two components

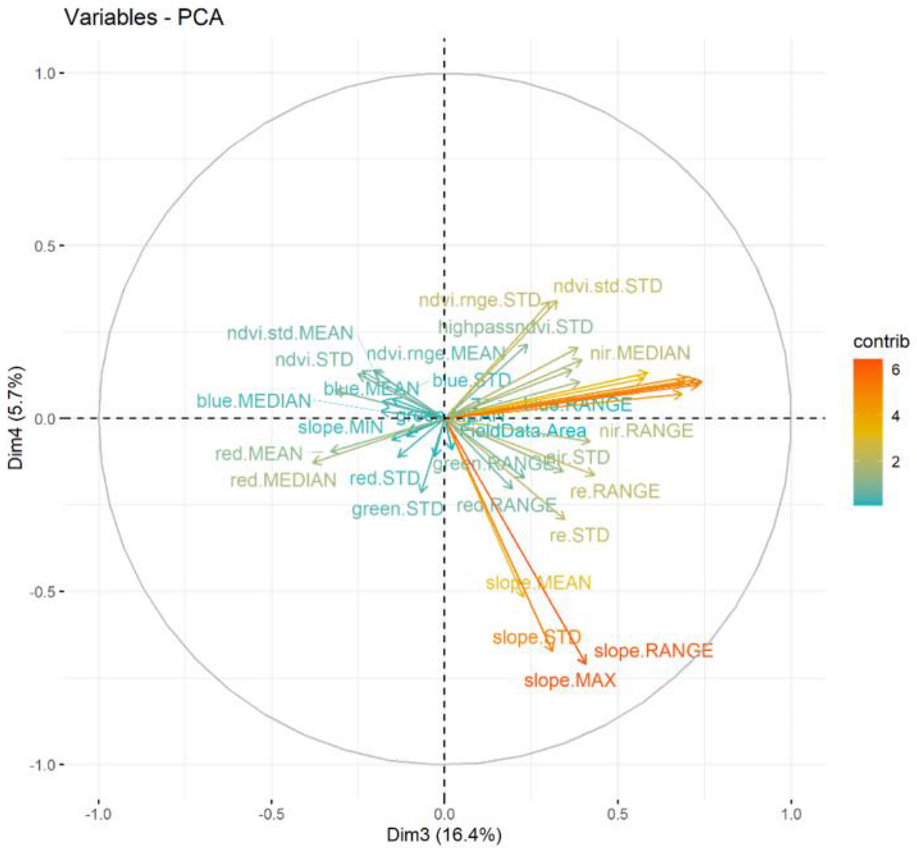


Figure 46. Loading plot of third and fourth components

6.4 Iso-Cluster Unsupervised Classification

This section shows the percent area of each class from the unsupervised classification rasters per LPIS field unit averaged per each grazing score group.

6.4a 3 Class

Using intuitive observation of NDVI values on the subject site, it appeared that the degree to which the surface was vegetated could be classified into three groups. These hypothetical three groups are non-vegetated surface (e.g. limestone, gravel drives, developed areas), intensively grazed grasslands and semi-natural priority habitat. Upon initial observation of the resulting raster, the Iso-Cluster Unsupervised Classification appeared to agree with the hypothesis.

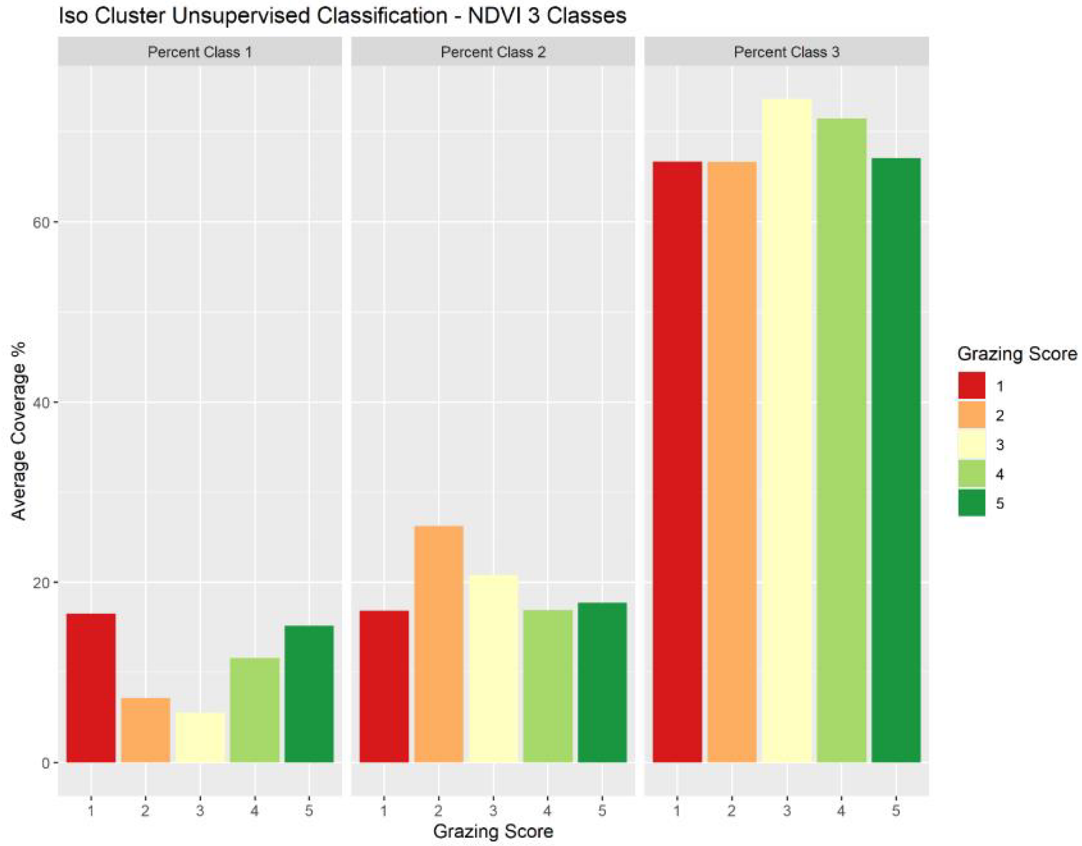


Figure 47. Average percent coverage of each classification class per grazing score

6.4b 5 Class

The 5 Class unsupervised classification raster was built to match the 5 grazing scores used by Caomhnú Árann. The Iso-Cluster Unsupervised Classification tool in ArcMap was unable to cluster 5 classes that satisfied the minimum number of pixels threshold of 20 with the NDVI as an alone input raster. So in turn, this unsupervised classification was built using the NDVI and Slope Raster. The bar plot in figure 48 displays the average tabulated area of each class within each grazing score group. The biggest outliers in this data visualization are the relative presences of Class 1 within Grazing Score groups one and five.

The Random Forest Model built with this input data included the original 317 observations of the newly created 5 variables. These variables were % Class 1 - % Class 5. For this model, with 5 input variables in total, the mtry value is 2. The minimum size for terminal nodes in this and all Random Forest Models in this thesis research is 1. The purpose of this model built with the tabulated areas of the unsupervised classification was to expose the Random Forest Classification to a different representation of the input data.

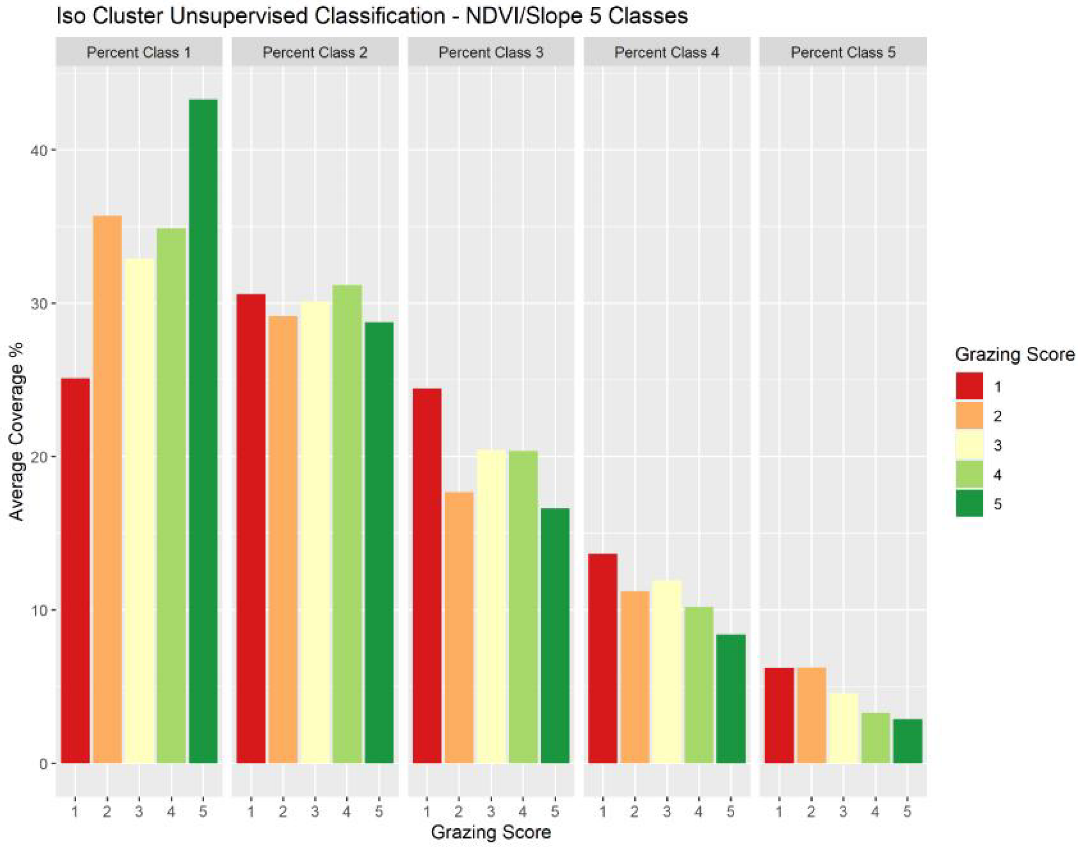


Figure 48. Average percent coverage of each classification class per grazing score

6.4c 10 Class

The 10 Class unsupervised classification raster was built to provide an extreme case of unsupervised classification. The Iso-Cluster Unsupervised Classification tool in ArcMap was unable to cluster 10 classes that satisfied the minimum number of pixels threshold of 20 with the NDVI as an alone input raster. So in turn, this unsupervised classification was built using the NDVI and Slope Raster, just the same as the 5 Class unsupervised raster. The bar plot in figure 49 displays the average tabulated area of each class within each grazing score group. The biggest outliers in this data visualization are the average relative presences of Classes 1, 2, & 3 fields scored five and one. Fields scored five on average also display a higher percent presence of class 5, 6, 7, & 8. The distinction between Grazing score groups one and five shows potential for this data to have success in the Random Forest Classification.

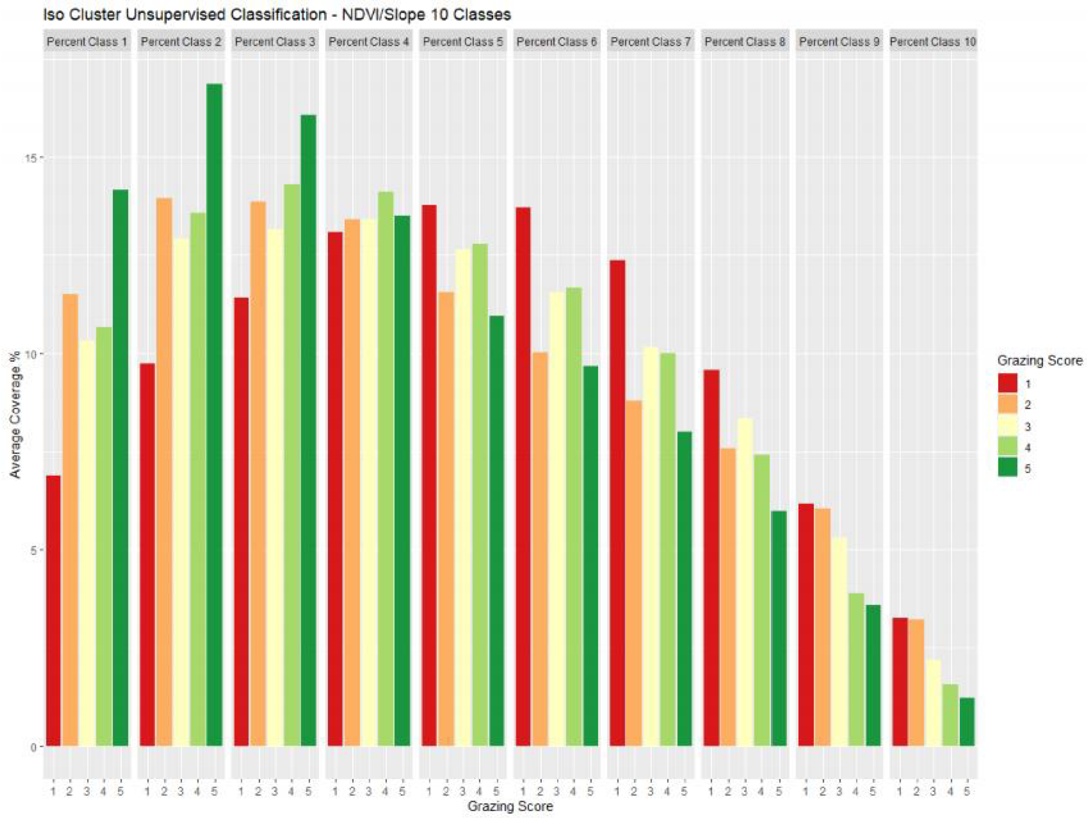


Figure 49. Average percent coverage of each classification class per grazing score

6.5 Random Forest Classification

Amongst the 6 supervised classification methods used, the model built from the first 10 dimensions of the Principal Component Analysis proved to be the most successful in predicting the grazing scores. It also possessed the highest Balanced Accuracy of any single class (class 5). This chapter will include detail the results of the confusion matrix and variable importance for each Random Forest Model. All confusion matrices shown in this chapter were constructed using the caret package (Kuhn 2022) in R and variable importance bar plots were built using the ggplot2 package in R.

6.5a PCA Random Forest Model

Table 5. PCA RF Model Confusion Matrix

PCA Random Forest Model

		Reference				
		1	2	3	4	5
Prediction	1	22	10	7	7	3
	2	22	75	26	10	7
	3	0	0	2	3	2
	4	3	3	2	4	2
	5	5	3	10	18	71

Overall Statistics

Accuracy :	0.5189
95% CI :	(0.4923, 0.6046)
No Information Rate	0.2871
P-Value [Acc > NIR] :	2.20E-16
Kappa :	0.3971
Mcnemar's Test P-Value :	8.55E-10

Statistics by Class

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.4231	0.8242	0.042553	0.09524	0.8353
Specificity	0.8981	0.7124	0.981481	0.96364	0.8418
Pos Pred Value	0.449	0.5357	0.285714	0.28571	0.6636
Neg Pred Value	0.8881	0.9096	0.854839	0.87459	0.9333
Prevalence	0.164	0.2871	0.148265	0.13249	0.2681
Detection Rate	0.0694	0.2366	0.006309	0.01262	0.224
Detection Prevalence	0.1546	0.4416	0.022082	0.04416	0.3375
Balanced Accuracy	0.6606	0.7683	0.512017	0.52944	0.8401

This model possessed the lowest OOB prediction error % and highest balanced accuracy for any individual Grazing Score class amongst the 6 random forest models. While its overall predicted accuracy was only .5489, its balanced accuracy for fields scored five was .8401. This may be due to the prevalence of Class Five amongst the data set. The same can also be said about Class Two. Classes Three and Four had remarkably low Sensitivity across all models. The low sensitivity of Classes Three and Four could be explained by the landscape characteristics of these classes. It should be noted although the grazing scores are ordinally ranked, they can not be considered interval. The grazing score has been designed to fit the specific

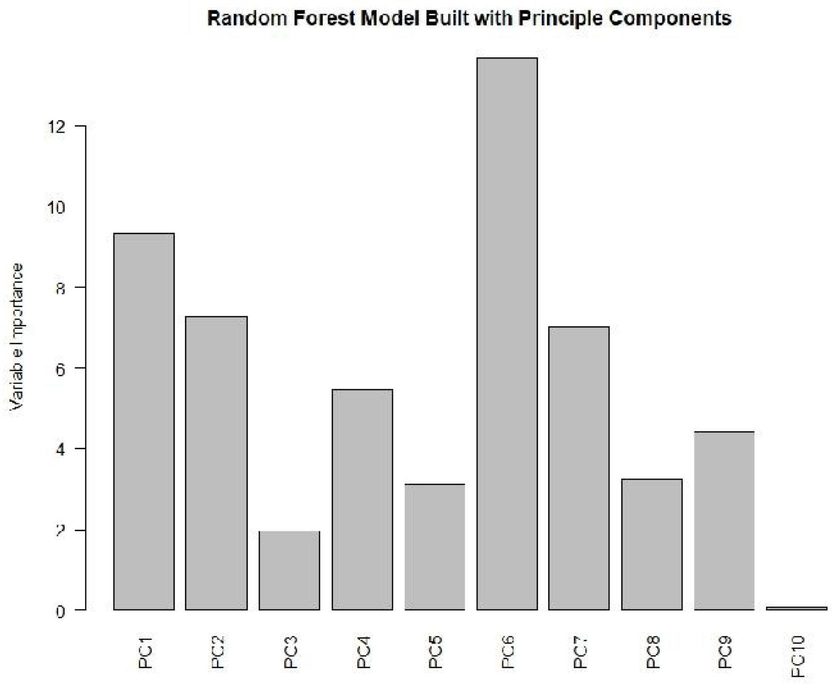


Figure 50. Variable Importance of PCA RF predictors

landscape phenomenon on the island. Threes and Fours have the potential to be Fives if they were grazed properly, and it is likely that they often exhibit similar mosaics of exposed limestone, thus resulting in similar NDVI values. In this Random Forest Model, 55.32% of the referenced "Threes" and 23.81% of the referenced "Fours" were predicted to be scored a Two. In line with this rhetoric is the percentage of Referenced Threes and Fours being predicted as having been scored Five (Three – 21.28% and Four-42.86%). In fact, all results in each of the 6 confusion matrices are comparable relative to themselves.

The PCA-built Random Forest returned some puzzling results in the form of variable importance as seen in figure 50. Most of the variability within the data set was projected onto the initial components, but we see a strong amount of variable importance within the random forest model from Principal Component 6. Figure 51 shows the constitution of the Dimensions 1 & 6 when plotted together. All variables showing a strong contribution also possess a strong positive loading on Dimension 1. The variable importance in the small and large random forest models will provide more insight into relevant data.

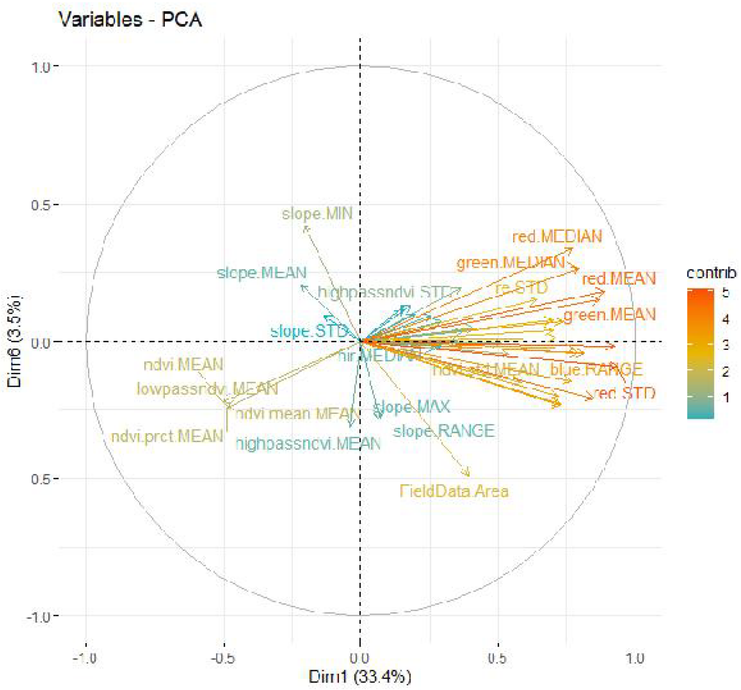


Figure 51. Loading plot of first and sixth components

6.5b Large Random Forest Model

Table 6. Large Model Confusion Matrix

Large Random Forest Model

		Reference				
		1	2	3	4	5
Prediction	1	15	4	2	4	2
	2	24	73	27	13	5
	3	7	8	2	5	3
	4	0	1	1	2	5
	5	6	5	15	18	70

Overall Statistics

Accuracy :	0.511
95% CI :	(0.4546, 0.5673)
No Information Rate	0.2871
P-Value [Acc > NIR] :	2.20E-16
Kappa :	0.3442
Mcnemar's Test P-Value :	1.74E-09

Statistics by Class

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.28846	0.8022	0.042553	0.047619	0.8235
Specificity	0.95172	0.6947	0.914815	0.974545	0.8103
Pos Pred Value	0.55556	0.5141	0.08	0.222222	0.614
Neg Pred Value	0.87241	0.8971	0.84589	0.87013	0.9261
Prevalence	0.16404	0.2871	0.148265	0.132492	0.2681
Detection Rate	0.04732	0.2303	0.006309	0.006309	0.2208
Detection Prevalence	0.08517	0.4479	0.078864	0.028391	0.3596
Balanced Accuracy	0.62159	0.7484	0.478684	0.511082	0.8169

The large model showed comparable ratios between the balanced accuracy of each of the 5 classes. And once again, incredibly low false or true prediction of classes 3 and 4. The results from the Correlation Matrix show the high collinearity between multiple variables within the original data set used to build the Large Model. By analyzing the variable importance of this model (figure 52) and the contribution of variables to the principal components, a smaller more selective model should yield a higher prediction accuracy. The dominating variable importance of variables derived from the Blue band is of concern because of issues with the data discussed in chapter XX.

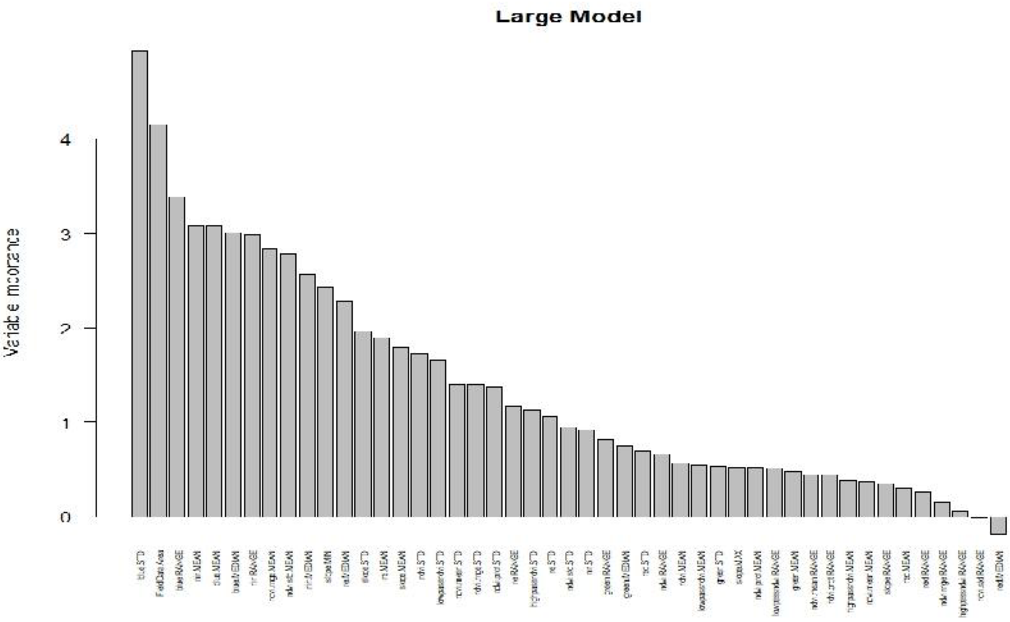


Figure 52. Variable Importance of Large Model predictors

6.5c Small Model

Table 7. Small Model Confusion Matrix

Small Random Forest Model						
		Reference				
		1	2	3	4	5
Prediction	1	15	6	1	7	4
	2	27	73	32	12	8
	3	3	2	1	0	0
	4	1	0	0	3	2
	5	6	10	13	20	71
Overall Statistics						
Accuracy :	0.5142					
95% CI :	(0.4577, 0.5704)					
No Information Rate	0.2871					
P-Value [Acc > NIR] :	2.20E-16					
Kappa :	0.3421					
Mcnemar's Test P-Value :	N/A					
Statistics by Class						
	Class: 1	**Class: 2**	**Class: 3**	**Class: 4**	**Class: 5**	
Sensitivity	0.28846	0.8022	0.021277	0.071429	0.8353	
Specificity	0.93208	0.6504	0.981481	0.989091	0.7888	
Pos Pred Value	0.45455	0.4803	0.166667	0.5	0.5917	
Neg Pred Value	0.86972	0.8909	0.85209	0.874598	0.9289	
Prevalence	0.16404	0.2871	0.148265	0.132492	0.2681	
Detection Rate	0.04732	0.2303	0.003155	0.009464	0.224	
Detection Prevalence	0.1041	0.4795	0.018927	0.018927	0.3785	
Balanced Accuracy	0.61027	0.7263	0.501379	0.53026	0.812	

The small model showed comparable ratios between the balanced accuracy of each of the 5 classes. And the lowest false or true prediction of classes 3 and 4. By analyzing the variable importance of this model (figure 53) the only observable trend is the low variable importance of variables representing the Standard Deviation zonal statistic. This model may have a slightly higher overall prediction accuracy than the Large Model but it is still not conclusive.

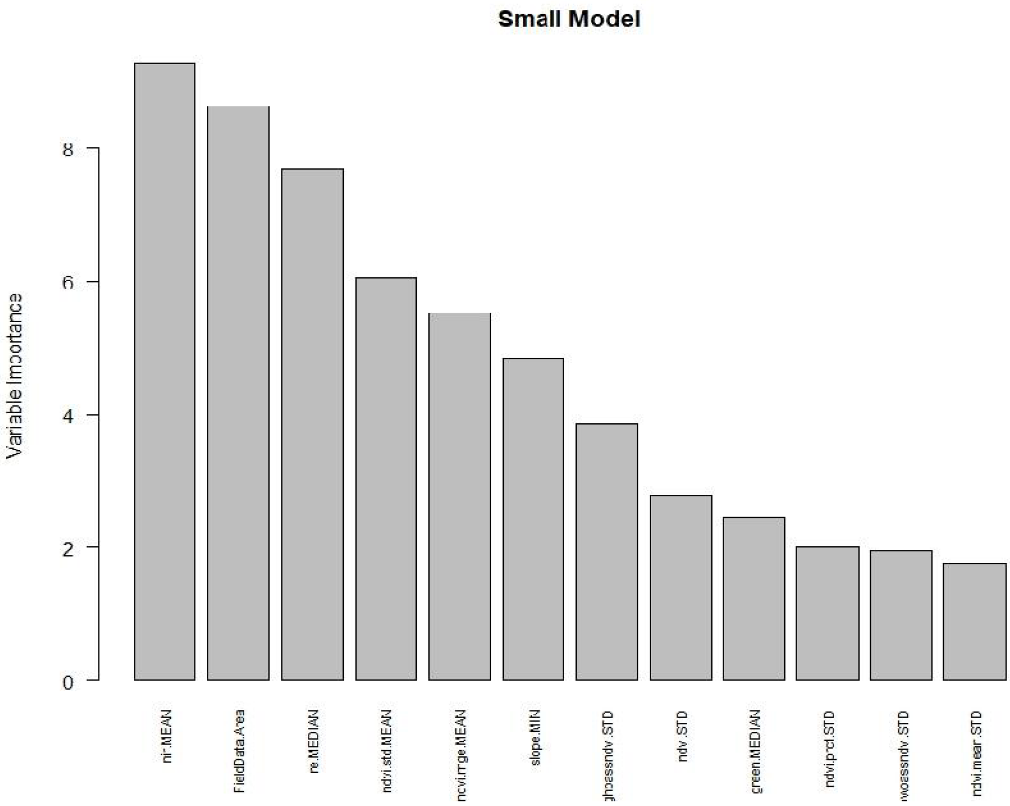


Figure 53. Variable Importance of Small Model predictors

6.5d Tabulated Area of Unsupervised Classification Rasters

6.5d.1 3 Class Model

Table 8. 3 Class tab. area RF Confusion Matrix

3 Class Iso-Cluster Unsupervised Classification Random Forest Model

		Reference				
		1	2	3	4	5
Prediction	1	8	6	8	2	6
	2	21	49	24	10	19
	3	5	13	3	5	3
	4	2	3	3	3	11
	5	16	20	9	22	46

Overall Statistics

Accuracy :	0.3438
95% CI :	(0.2917, 0.399)
No Information Rate	0.2871
P-Value [Acc > NIR] :	0.01604
Kappa :	0.1306
Mcnemar's Test P-Value :	0.00194

Statistics by Class

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.15385	0.5385	0.06383	0.07143	0.5412
Specificity	0.91698	0.6726	0.9037	0.93091	0.7112
Pos Pred Value	0.26667	0.3984	0.10345	0.13636	0.4071
Neg Pred Value	0.84669	0.7835	0.84722	0.8678	0.8088
Prevalence	0.16404	0.2871	0.14827	0.13249	0.2681
Detection Rate	0.02524	0.1546	0.00946	0.00946	0.1451
Detection Prevalence	0.09464	0.388	0.09148	0.0694	0.3565
Balanced Accuracy	0.53541	0.6055	0.48377	0.50117	0.6262

The Random Forest Model built using the tabulated area of the 3 class Iso-Cluster Unsupervised Classification raster produced the lowest prediction accuracy of all 6 models. It performed so poorly that the Balanced Accuracy of Class 2 and Class 5 which performed moderately well in the previous 3 models are nearly on par with the balanced accuracy of the other three classes. Noticeable is the greater number of predictions of Class 3 and 4 compared to the other models. This is also true in the Random Forest Model built using the tabulated area of the 5 class Iso-Cluster Unsupervised Classification raster. Similar to the results of the previous models, a plurality of the fields referenced as 3 were predicted as a 2, and a plurality of the fields referenced as 4 were predicted as a 5.

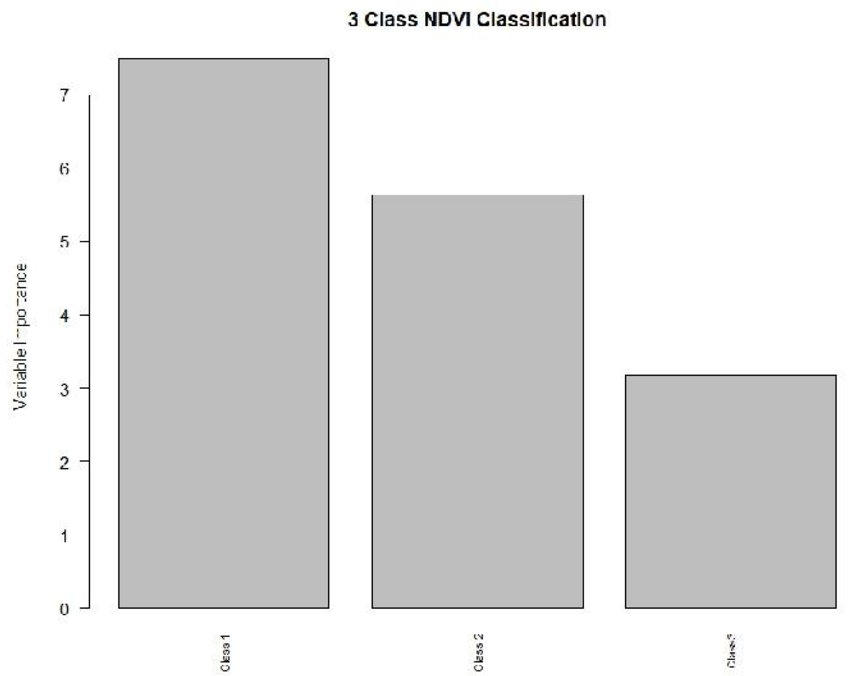


Figure 54. Variable Importance of 3 class RF model

6.5d.2 5 & 10 Class Models

The performance of the model built using the 10 class Iso-Cluster Unsupervised Classification raster more closely resembled the results of the PCA, Large, and Small models than the other Iso-Cluster Unsupervised classification models. See figures XX for the confusion matrices for the 5 and 10 class models.

Table 9. 5 Class tab. area RF Confusion Matrix

5 Class Iso-Cluster Unsupervised Classification Random Forest Model

		Reference				
		1	2	3	4	5
Prediction	1	21	15	12	5	6
	2	13	55	19	9	18
	3	2	1	1	4	4
	4	8	6	3	7	12
	5	8	14	12	17	45

Overall Statistics

Accuracy :	0.4069
95% CI :	(0.3524, 0.4633)
No Information Rate	0.2871
P-Value [Acc > NIR] :	3.23E-06
Kappa :	0.2241
Mcnemar's Test P-Value :	0.000691

Statistics by Class

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.40385	0.6044	0.021277	0.16667	0.5291
Specificity	0.8566	0.7389	0.959259	0.89455	0.7802
Pos Pred Value	0.35593	0.4825	0.083333	0.19444	0.4688
Neg Pred Value	0.87984	0.8227	0.84918	0.87544	0.819
Prevalence	0.16404	0.2871	0.148265	0.13249	0.2681
Detection Rate	0.06625	0.1735	0.003155	0.02208	0.142
Detection Prevalence	0.18612	0.3596	0.037855	0.11356	0.3028
Balanced Accuracy	0.63022	0.6717	0.490268	0.53061	0.6548

Table 10. 10 Class tab. area RF Confusion Matrix

10 Class Iso-Cluster Unsupervised Classification Random Forest Model

		Reference				
		1	2	3	4	5
Prediction	1	15	7	10	5	8
	2	25	64	22	18	14
	3	3	3	3	2	1
	4	1	3	2	2	1
	5	8	14	10	15	61

Overall Statistics

Accuracy :	0.4574
95% CI :	(0.4016, 0.514)
No Information Rate	0.2871
P-Value [Acc > NIR] :	9.69E-11
Kappa :	0.2731
Mcnemar's Test P-Value :	2.03E-09

Statistics by Class

	Class: 1	Class: 2	Class: 3	Class: 4	Class: 5
Sensitivity	0.28846	0.7033	0.06383	0.047619	0.7176
Specificity	0.88679	0.6504	0.966667	0.974545	0.7974
Pos Pred Value	0.33333	0.4476	0.25	0.222222	0.5618
Neg Pred Value	0.86397	0.8448	0.855738	0.87013	0.8852
Prevalence	0.16404	0.2871	0.148265	0.132492	0.2681
Detection Rate	0.04732	0.2019	0.009464	0.006309	0.1924
Detection Prevalence	0.14196	0.4511	0.037855	0.028391	0.3407
Balanced Accuracy	0.58763	0.6769	0.515248	0.511082	0.7575

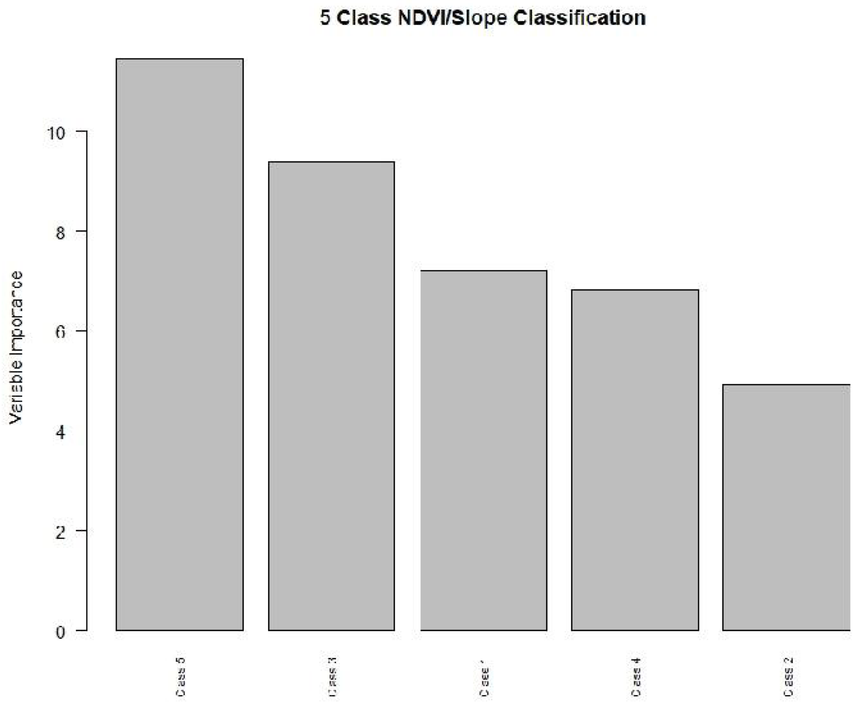


Figure 55. Variable Importance of 5 class RF model

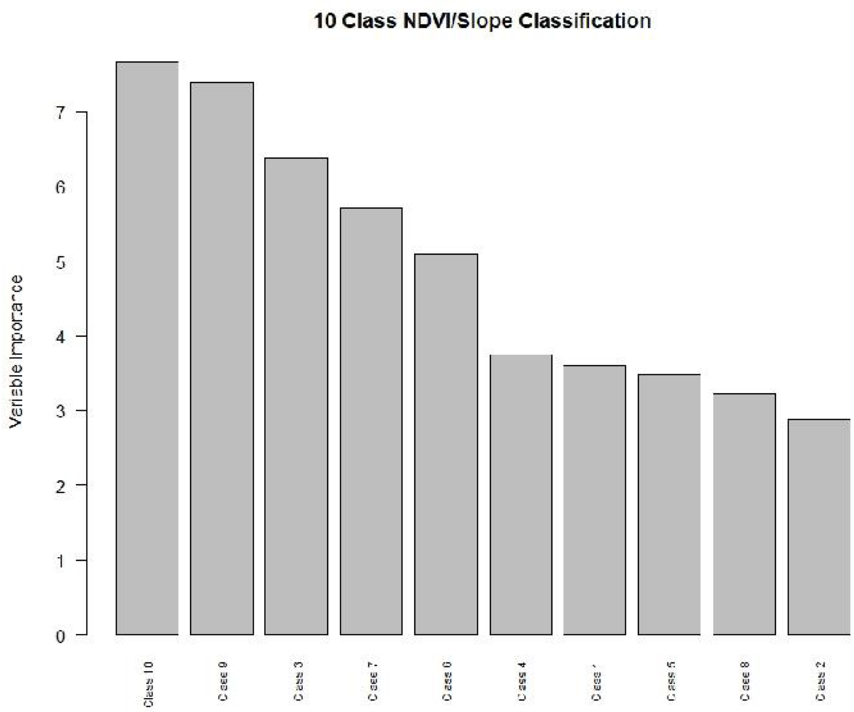


Figure 56. Variable Importance of 10 class RF model

Chapter 7

Discussion

Upon investigation of the input data, some flaws were discovered. Across all sites, there appears to be a striping issue. This can be a result of the sun's position in the sky, the chosen gimble angle, or the cloud cover at the time of the survey. The original images acquired by GeoAerospace were not provided for this thesis and that limited the ability to troubleshoot this issue. The result, however, was an Orthomosaic with 30m wide stripping running north-south. The striping is realized by differentiating pixel values for otherwise similar objects. It can be visualized quite well in the NDVI, where along gravel roads, one can witness NDVI value jumping $\approx$.4 every roughly 30 m. This thesis suggests any future use of this data set would first attempt to correct the atmospheric disturbances using a Fast Fourier Transformation, Flat Field Correction, or Internal Average Relative Reflectance (IARR) (Chen 2017) (Kokka 2019) (Ben-Dor 1994) (THOR). Figures XX below displays the raster statistics for the multispectral imagery captured in each of the 3 sites.

Table 11. Statistics of multispectral bands per site

Blue Values			
Site	1	2	3
Min	154	0	0
Max	7659	21103	8025
Mean	1622.25	2082.87	572.874
SD	982.218	1544.9	378.057

Green Values			
Site	1	2	3
Min	348	0	0
Max	9635	28576	21659
Mean	2798.08	3126.16	2146.9
SD	953.735	1548.25	895.66

Red Values			
Site	1	2	3
Min	53	0	0
Max	10530	35988	28179
Mean	2201.34	2588.16	1773.52
SD	1326.68	1973.42	1203.98

Red Edge Values			
Site	1	2	3
Min	21	0	0
Max	21096	57722	56665
Mean	9077.87	10247.7	6808.21
SD	1981.24	3987.97	2668.25

Near-Infrared Values			
Site	1	2	3
Min	1	0	0
Max	32386	65535	64755
Mean	13812.5	15797.5	9409
SD	3477.15	6614.52	3875.24

Another issue with the provided multispectral data was the significantly low values from the Blue band in the site known as Mór 3 (NE site). Due to the limited time and resources associated with this research, and the difficulty of communicating by email, this thesis cannot provide an answer for this anomaly. It is for this reason, however, that no variables sourced from the Blue band were included in the “Small” model. This research could have worked off of the theories outlined in Strong’s paper much more, specifically the usefulness of the ENDVI index, if the values of the Blue band were not in question.

This research differs from most grassland classification attempts in that it attempted to classify pre-existing field units, often made up of multiple pastures, by a pre-existing scoring structure. Most classification attempts (Xu 2019) (Svoboda 2022) are at the pixel level, not polygon.

Chapter 8

Conclusion

While a supervised classification model such as a Random Forest was not proven to accurately classify any grazing score, there was promise in the ability to classify fields scored 2 and 5. If the large sample size of these two groups is affecting the model's ability to classify these fields, then there is also promise in the prospect of the model being more successful with a larger data set or perhaps temporal data. This thesis if anything, gives Caomhnú Árann a starting point for their prospective remote sensing endeavors.

.

Bibliography

Reports

Browne, Amanda. (2020) "Should Farmers Score Their Own Fields in Results Based Programmes?" *NRN EIP-AGRI BLOG*, EIP-AGRI, https://www.caomhnuaranneip.ie/_files/ugd/0ca2f5_4e0a35e3e2ac45a690e3ffa004bbc7fd.pdf?index=true.

Caomhnú Árann, Inis Oírr, Ireland, 2018, *Caomhnú Árann Field scoring system for assessing habitat condition,* https://www.caomhnuaranneip.ie/_files/ugd/0ca2f5_ed02b524086147bd898d802d2243e259.pdf?index=true.

Caomhnú Árann, Inis Oírr, Ireland, 2019, *Caomhnú Árann EIP Project Annual Report*.

Caomhnú Árann, Inis Oírr, Ireland, 2020b, *Newsletter Issue 3*. https://www.caomhnuaranneip.ie/_files/ugd/0ca2f5_6bb8b23674c24f2ead02043ca771c73c.pdf

GeoAerospace.(2019) n.d., *Caomhnú Árann - Satellite Grazing Scores Classification*.

Kuhn, Max. "Caret: Classification and Regression Training." *Classification and Regression Training*, 11 Mar. 2022, https://cran.r-project.org/web/packages/caret/caret.pdf.

Wright, Marvin N. "A Fast Implementation of Random Forests." *CRAN*, Comprehensive R Archive Network (CRAN), 14 July 2021, https://cran.r-project.org/web/packages/ranger/.

Journals

Bassett, J. A., and T. G. F. Curtis. "The Nature and Occurrence of Sand-Dune Machair in Ireland." Proceedings of the Royal Irish Academy. Section B: Biological, Geological, and Chemical Science, vol. 85B, Royal Irish Academy, 1985, pp. 1–20, http://www.jstor.org/stable/20494442.

Ben-Dor, E., and F. A. Kruse. "The Relationship Between the Size of Spatial Subsets of GER 63 Channel Scanner Data and the Quality of the Internal Average Relative Reflectance (IARR) Atmospheric Correction Technique." *International Journal of Remote Sensing*, vol. 15, no. 3, Feb. 1994, pp. 683–90, https://doi.org/10.1080/01431169408954107.

Chen, Yong, et al. “Stripe Noise Removal of Remote Sensing Images by Total Variation Regularization and Group Sparsity Constraint.” *Remote Sensing*, vol. 9, no. 6, 2017, p. 559., https://doi.org/10.3390/rs9060559.

He, Yuhong & Guo, Xulin & Wilmshurst, John. (2006). “Studying mixed grassland ecosystems I: Suitable hyperspectral vegetation indices.” *Canadian Journal of Remote Sensing*. 32. 98-107. 10.5589/m06-009.

Jackson, Donald A. “Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches.” Ecology, vol. 74, no. 8, Ecological Society of America, 1993, pp. 2204–14, https://doi.org/10.2307/1939574.

López, J (2019), Drones for Conservation in Protected Areas: Present and Future. *Drones*. 2019; 3(1):10. https://doi.org/10.3390/drones3010010

Kokka, Alexander, et al. “Flat-Field Calibration Method for Hyperspectral Frame Cameras.” *Metrologia*, vol. 56, no. 5, 2019, p. 055001., https://doi.org/10.1088/1681-7575/ab3261.

Komárek, Jan, et al. “The Potential of Unmanned Aerial Systems: A Tool towards Precision Classification of Hard-to-Distinguish Vegetation Types?” *International Journal of Applied Earth Observation and Geoinformation*, vol. 71, 2018, pp. 9–19., https://doi.org/10.1016/j.jag.2018.05.003.

Lemenkova, Polina. “ISO Cluster Classifier by Arcgis for Unsupervised Classification of the Landsat TM Image of Reykjavík.” *Bulletin of Natural Sciences Research*, vol. 11, no. 1, 2021, pp. 29–37., https://doi.org/10.5937/bnsr11-30488.

McGurn, Patrick, and Bernard McKay (2020b). “Nature and Agriculture: A Future of Divergence or Convergence?” Biology and Environment: Proceedings of the Royal Irish Academy, vol. 120B, no. 2, Royal Irish Academy, 2020, pp. 83–90, https://doi.org/10.3318/bioe.2020.12.

McGurn P. (2017), et al. “Semi-natural grasslands on the Aran Islands, Ireland: ecologically rich, economically poor.” *Grassland Science in Europe*, vol. 22, 7-10 May 2017., pp. 197–199.

Molloy, Karen, and Michael O’Connell. “Holocene Vegetation and Land-Use Dynamics in the Karstic Environment of Inis Oırr, Aran Islands, Western Ireland: Pollen Analytical Evidence Evaluated in Light of the Archaeological Record.” Quaternary International 113.1 (2004): 41–64. Web.

O’Rourke, Eileen. “Socio-Natural Interaction and Landscape Dynamics in the Burren, Ireland.” *Landscape and Urban Planning*, vol. 70, no. 1-2, 2005, pp. 69–83., https://doi.org/10.1016/j.landurbplan.2003.10.015.

Rensburg, Thomas & Kelley, Hugh & Yadav, Lava. (2009). “Farming for Conservation of the Upland Landscape and Biodiversity in the Burren.” National University of Ireland Galway, Department of Economics, Working Papers.

Schuster, Christian, et al. (2012) “Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data.” *International Journal of Remote Sensing*, 33:17, 5583-5599, DOI: 10.1080/01431161.2012.666812

Strong CJ, Burnside NG, Llewellyn D (2017) The potential of small-Unmanned Aircraft Systems for the rapid detection of threatened unimproved grassland communities using an Enhanced Normalized Difference Vegetation Index. PLOS ONE 12(10): e0186193. https://doi.org/10.1371/journal.pone.0186193

Sun, Zheng-guo, et al. “Classification and Net Primary Productivity of the Southern China's Grasslands Ecosystem Based on Improved Comprehensive and Sequential Classification System (CSCS) Approach.” *Journal of Integrative Agriculture*, vol. 13, no. 4, 2014, pp. 893–903., https://doi.org/10.1016/s2095-3119(13)60415-3.

Svoboda, Jan, et al. “Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia.” *Remote Sensing*, vol. 14, no. 5, 2022, https://doi.org/10.3390/rs14051189.

Taddia, Y., et al. “Using DJI Phantom 4 RTK Drone for Topographic Mapping of Coastal Areas.” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W13, 2019, pp. 625–630., https://doi.org/10.5194/isprs-archives-xlii-2-w13-625-2019.

Xu, Dawei et al. (2019) "The Classification of Grassland Types Based on Object-Based Image Analysis with Multisource Data," Rangeland Ecology and Management, 72(2), 318-326, (5 March 2019)

Y. Feng and J. Wang, "GPS RTK Performance Characteristics and Analysis," *Positioning*, Vol. 1 No. 13, 2008, pp. -.

Zhao, Yifan, et al. “Classification of Zambian Grasslands Using Random Forest Feature Importance Selection during the Optimal Phenological Period.” Ecological Indicators, vol. 135, Feb. 2022, p. 108529, https://doi.org/10.1016/j.ecolind.2021.108529.

Books

Colgan, Nathaniel. “Notes on the Flora of the Aran Islands.” *The Irish Naturalist*, vol. 2, no. 3, Irish Naturalists’ Journal Ltd., 1893, pp. 75–78, http://www.jstor.org/stable/25520348.

Drew, D., 1997. The Burren, County Clare. In: Aalen, F.H.A., Whelan, K., Stout, M. , Atlas of the Irish Rural Landscape. Cork University Press, Cork, pp. 287298.

Dunford, Brendan, and Sharon Parr. "Farming for Conservation in the Burren ." *Farming for Nature The Role of Results-Based Payments*, Teagasc , Wexford, Ireland, 2020, pp. 56–105.

Jackson, J. Edward. *A User's Guide to Principal Components*. Wiley-Interscience, 2003.

McGurn, Patrick. et al. "Farming for Conservation on the Aran Islands." *Farming for Nature The Role of Results-Based Payments*, Teagasc , Wexford, Ireland, 2020, pp. 108–147.

Robinson, Tim. *Stones of Aran: Pilgrimage*. Faber and Faber, 2008.

Webpagess

Anonymous. "Caomhnú Árann Managing the Habitats of the Aran Islands to Maximise Their Agricultural & Ecological Output." *EIP-AGRI - European Commission*, 28 Aug. 2019, https://ec.europa.eu/eip/agriculture/en/find-connect/projects/caomhn%C3%BA-%C3%A1rann-managing-habitats-aran-islands.

Self, C.A. (1998) "The Caves of the Aran Islands, County Galway, Ireland - UBSS." *University of Bristol Spelaeological Society*, "unpublished" http://www.ubss.org.uk/resources/proceedings/vol21/UBSS_Proc_21_2_159-173.pdf.

"Land Parcel Identification System (LPIS)." *Government of Ireland*, 23 Oct. 2019, https://www.gov.ie/en/service/1eb4d-land-parcel-identification-system-lpis/.

LDP LLC. "LDP LLC." *ENDVI*, https://maxmax.com/maincamerapage/remote-sensing/enhanced-normalized-difference-vegetation-index.

"THOR Atmospheric Correction." *l3harrisgeospatial*, https://www.l3harrisgeospatial.com/docs/thoratmosphericcorrection.html.

Correspondence

Burnside, Niall G. Received by Patrick McGurn, *Drone Technology NDVI*, 7 Jan. 2020.

Fearghus Foyle. Received by Patrick McGurn, *Edited Shapefile*, 28 March 2022.

Data

“© OpenStreetMap contributors.” “Counties - Shapefile(.Shp).”