THE UNIVERSITY of ADELAIDE

Improved hyperspectral classification of vegetation through generative deep learning models

Thesis submitted for the degree of

Doctor of Philosophy

Andrew James Hennessy

School of Biological Sciences

The University of Adelaide

March 2021

Intentionally left blank

ABSTRACT

Early studies into hyperspectral reflectance demonstrated that the spectra of different plants have the potential for taxonomic discrimination and classification, though this came with the caveat that misidentification was a frequent impediment as a result of small sample sizes, inter-class similarity and intra-class variability.

The aim of this thesis was to develop methods of improving the ratio between intra and inter-class variability in hyperspectral vegetation spectra, and ultimately increasing classification accuracy, reliability and generalisability.

This was addressed in three ways: (1) reviewing the hyperspectral classification literature of the past two decades, while also performing pre-processing and classification trials on a vegetation hyperspectral dataset, (2) increasing the number and distribution of classifier training samples through data augmentation, and (3) improving the intra/inter-class relationship with deep generative sample transformation.

For objective one the last two decades of hyperspectral vegetation classification literature was systematically reviewed, specifically focusing on waveband/feature selection. Additionally, waveband selection trials were performed on a curated hyperspectral dataset in order to test the findings of the review. Both the review and waveband selection trials indicated that all characteristics of hyperspectral plant studies influence the wavebands selected for classification. However, the considerable variability in waveband selection caused by the chosen feature selection method effectively masked analysis of any variability in waveband selection caused by other aspects of the studies in the review. For this reason caution is suggested in relying upon waveband recommendations from the literature to guide waveband selections or classifications for new plant discrimination applications. As such recommendations appear to be weakly generalizable between studies.

The data augmentation performed for objective two was realised through the use of a generative adversarial network (GAN), a type of generative deep learning model that could produce realistic synthetic hyperspectral vegetation data. After being trained on

vegetation spectra the GAN was able to generate synthetic samples that visually matched the training spectra as well as statistically matched the distribution of each vegetation class in the training data. Creation of an augmented dataset consisting of synthetic and original samples produced training datasets with far greater sample sizes. Under almost all circumstances increases in classification accuracy of multiple classifiers was seen following their training with the augmented dataset.

Objective three expanded upon the data augmentation abilities of the GAN used in objective two, introducing the ability to replicate sample spectra whilst transforming them based upon the learned features of the other classes in the study. These transformations were performed to manipulate the intra/inter-class relationships in a desired manner. This was performed on both the training and evaluation subsets of a hyperspectral vegetation dataset producing *n* transformed replicates of every sample where *n* is the number of classes in the study. Training and then evaluating the accuracy of each of these transformed subsets with multiple classification methods produced accuracies significantly higher than that of the original dataset. This significant increase in accuracy was then further improved following the ensembling of the *n* classification results. Visualisation of the samples used in the ensembled classification following projection to 2d space showed samples to be tightly clustered by class, indicating the successful reduction of intra-class variance as well the reduction in inter-class overlap.

This thesis represents a significant step towards eliminating the lack of generalisability and transferability of vegetation classification models resulting from the pronounced effects of intra-class variance and inter-class similarity. It presents the opportunity for remote sensing practitioners to deploy their classification models to greater spatial and temporal extents whilst giving extra utility to hyperspectral samples contained within spectral libraries.

DECLARATION

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Andrew Hennessy

March 2021

PUBLICATIONS ARISING FROM THIS THESIS

Hennessy, A., K. Clarke and M. Lewis (2020). "Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability." *Remote Sensing* 12(1): 113

Hennessy, A., K. Clarke and M. Lewis (2021). "Generative adversarial network synthesis of hyperspectral vegetation data." *Remote Sensing* 13(12): 2243.

Hennessy, A., K. Clarke and M. Lewis (*Submitted*). “Generative adversarial network-based style transfer for data augmentation of hyperspectral vegetation data.” *Remote Sensing*.

ACKNOWLEDGMENTS

Firstly, and most importantly, I would like to thank my fantastic supervisors Megan Lewis and Ken Clarke. Thank you for taking me on as a student and thank you for all of your support and guidance over the last five’ish years, and here’s to hoping for many more years of collaboration and friendship. I’d like to thank the Spatial Science Group and all other denizens of Oliphant Level 3. Thank you for your support, for all the morning and afternoon teas, the unibar beers, and sneaky office whisky.

To my parents and family, thank you for your love and preparing me for this journey and the continued support along the way. I am so glad that I took on this adventure and thank you so much for supporting me during this part of my life.

Lastly, I’d like to thank all of my friends old and new who have supported me during this time and who continue to support me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: Introduction

1.1 Background

The need to map and take inventory of vegetation around the globe has become essential for the protection of our natural floristic resources. These resources offer more beyond their intrinsic ecological value: they provide habitats for wildlife, affect soil condition and water availability, and have considerable economic, scenic and societal value. Factors such as overgrazing, land clearing, invasive species, and poor land management all threaten the condition, distribution and biodiversity of the world's vegetation. As well as these impacts, the effects of climate change are ever increasing, resulting in migrations across latitude and elevation gradients, phenotypic responses, and changes in biodiversity. Due to the scale required to map and monitor the world's vegetation, fast, wide ranging, repeatable, objective methods that provide results that can be quickly and easily shared and analysed are required. One such potential method is the use of reflectance spectra recorded from leaves, entire plant canopies or plant communities. Reflectance spectroscopy in optical wavelengths of the electromagnetic spectrum fulfils these requirements, producing objective, digital measurements that can be easily shared and quickly analysed with semi-automated procedures in a repeatable and objective manner. Reflectance spectra of plants have the potential to classify samples into varying taxonomic and functional groups, while also providing information on general condition, including productivity, stress, and lifecycle phase. Early studies into spectral reflectance demonstrated that the spectra of different plants have the potential for discrimination and classification, though this came with the caveat that misidentification was a real possibility as a result of low sample size, spectral mixing, interspecies similarity and within-species variability (Price 1994).

The degree of taxonomic and functional discrimination possible is highly dependent on the taxa being classified, the key factor being the degree of dissimilarity between the spectra of all taxa or groups included in the study. This is often the case when classifying and mapping invasive species, as the invasive plant regularly exhibits differences in growth form or phenological cycle in comparison to native or agricultural species (Underwood, Ustin et al. 2003, Bolch, Santos et al. 2020, Papp, van Leeuwen et al. 2021). However, due to the convergent nature of many plant traits, plants that are closely related, or unrelated though growing under the same environmental conditions often

express very similar traits, and therefore spectral reflectance. Despite this, many studies have succeeded in classifying plants with high accuracy, even when closely related or physically similar. However, a decrease in classification accuracy is evident in studies including multiple species of the same genus (Pu 2009), with a greater decrease seen for genera that share particular environmental adaptations (Goodwin, Turner et al. 2005). Although classification between closely related taxa is possible, Cavender-Bares, Meireles et al. (2016) have demonstrated a substantial increase in classification accuracy when expanding from population to species to clade levels. The methods used in studies that include both closely related and unrelated plants may also decrease accuracy in discriminating the closely related species. For example, Peerbhay, Mutanga et al. (2013) noted that the dimensionality reduction technique and subsequent statistical selection of maximally discriminative bands may favour discrimination between unrelated taxa, with the minor variations between closely related species being discarded.

Solar radiation follows the distribution of a black body at 5800 K from 200 to 2500 nm, with a peak at ~500 nm. It is for this reason that photosynthetic pigments absorb in the region of maximum solar irradiance of 400 -700 nm, as well as being the reason that reflectance spectroscopy generally covers the region of ~350 – 2500 nm. While definitions are often dependant on context the wavelengths tend to be broken into three spectral regions: visible (~400 – ~700 nm), near infrared (~700 - ~1400 nm), and shortwave infrared (~1400 - ~2500 nm) (Asner 1998).

Reflection of plants in the visible wavelengths (400 – 700 nm) is dominated by absorptions from foliar pigments, with chlorophyll a and b having the strongest influence in this region. Absorption from less influential pigments such as carotenoids and anthocyanins are predominantly masked by that of chlorophyll, only becoming dominant during periods of senescence or stress (Curran 1989).

The red edge encompasses the region from the red reflectance minimum around 680 nm to the near infrared (NIR) shoulder at approximately 780 nm and indicates the sharp increase in reflectance from the visible (VIS) to the NIR plateau (Jacquemoud, Verhoef et al. 2009). The red edge region has been described as one of the most informative and frequently selected regions, its importance has been attributed to its correlation with

chlorophyll abundance, nitrogen concentration, water content and structural features such as Leaf Area Index (LAI) (Clark, Roberts et al. 2005, Dalponte, Bruzzone et al. 2009, Cho, Debba et al. 2010, Dalponte, Bruzzone et al. 2012). The NIR plateau (780 – 1327 nm) is a region of high reflectance resulting from scattering of photons within the leaf structure due to a change in the refractive index from liquid water to air within the inter-cellular spaces (Knipling 1970). High levels of intraspecific variability have been identified in the NIR related to leaf age, water and chlorophyll concentration, as well as herbivory, necrosis and epiphyll cover (Datt 2000, Clark, Roberts et al. 2005).

The shortwave infrared wavelengths (SWIR) can be divided into two distinct regions, the near SWIR (NSWIR) from 1350 – 1800 nm including the strong water absorption feature at 1350 – 1450 nm, and the far SWIR (FSWIR) from 1800 -2500 nm, including another strong water absorption feature from 1800 – 2000 nm. Reflectance from the SWIR is reported as being associated with water absorption (Datt 2000, Thenkabail, Enclona et al. 2004, Adam, Mutanga et al. 2012, Ferreira, Grondona et al. 2013, Thenkabail, Mariotto et al. 2013, Fassnacht, Neumann et al. 2014, Shang and Chisholm 2014, Ferreira, Zortea et al. 2016) or the weak harmonic and overtone absorptions from biochemicals such as lignin, starch, and cellulose (Thenkabail, Enclona et al. 2004, Wang, Xu et al. 2009, Jones, Coops et al. 2010, Ferreira, Grondona et al. 2013, Thenkabail, Mariotto et al. 2013, Alonzo, Bookhagen et al. 2014, Shang and Chisholm 2014, Lehmann, Große-Stoltenberg et al. 2015, Ferreira, Zortea et al. 2016).

Variability in plant spectral properties complicates supervised classifications, as there may be considerable differences between reference spectra used for training and the spectra or imagery targeted for classification. Sources of variability in the spectra fall within three categories: measurement, leaf level, and canopy level.

1.1.1 Measurement Variability

Spectral variability that occurs during measurement is dependent upon the method of capture, whether collected in the laboratory or field, at leaf, stem or canopy level, or remotely sensed at canopy scale. Each of these methods must contend with different illumination conditions as well as differing soil or background reflectance. Measurements at leaf scale reduce much of the variability inherent in other methods.

However, leaf level measurements may not be representative of the entire plant, as foliar reflectance will vary within a leaf, and between leaves of the same plant (Castro-Esau, Sánchez-Azofeifa et al. 2006, Tirado, St Dennis et al. 2020). Laboratory measurements allow for control and standardisation of both illumination and background reflectance, while field measurements have the option to control background reflectance, although generally rely upon natural solar illumination.

Variation in natural solar irradiance is the biggest variable for reflectance measurements, with sun angle relative to the sample having the greatest effect. Sun angle varies throughout the day and year and is dependent on the latitude of the observations. As well as solar angle, the amount of light incident on a surface can be varied by absorption and scattering of the incoming photons by atmospheric gases and dust particles as the light traverses the atmosphere. Different atmospheric gases absorb at specific wavelengths, the most significant of which are associated with water vapour. Almost complete absorption by water occurs at ~1400 and ~1900 nm, with smaller bands at ~900 and ~1100 nm (Salisbury 1998). Along with this ubiquitous atmospheric effect are intermittent atmospheric effects such as cloud cover, smoke, dust and pollution, all of which will influence measured reflectance and must be taken into consideration during spectral collection. To account for this variation in solar irradiance at the time of spectral measurement, the recording of a standard surface with known reflectance values can be used for the calibration of spectra from target samples (Held, Phinn et al. 2015). Additionally, post-processing methods such as physical atmospheric correction of fully shaded pixels in comparison to directly illuminated areas enables the aerosol optical thickness to be estimated and corrected (Schläpfer, Hueni et al. 2018).

Collection of imagery from airborne or satellite imagers must also contend with spatial distortions as well as increased atmospheric effects, as the reflected radiation must also traverse the atmosphere before being recorded. Image restoration procedures to correct for these variations include radiometric restoration (Guo, Wang et al. 1998), which attempts to remove or reduce spectral distortions, and geometric restoration that corrects for spatial distortions and registers the image to conventional mapping coordinates (Richards and Richards 1999, Richter, Schläpfer et al. 2011). Methods for atmospheric correction include empirical approaches that remove atmospheric effects

by dividing pixels by the average spectrum of a scene (Kruse 1988) or applying a transformation derived from an area of neutral spectral reflectance within the scene (Roberts, Yamaguchi et al. 1986). Other empirical methods require field measurements of at least one bright and one dark target within the scene allowing for the derivation of gain and offset curves (Curtiss 1987). Alternatively, radiative transfer models have been used to simulate the transmission spectrum of water vapour and other gases dominant in the 400 – 2500 nm region (Gao, Heidebrecht et al. 1993, Thompson, Natraj et al. 2018). Hybrid models have also been developed using both radiative transfer modelling and empirical field spectral measurements (Goetz, Boardman et al. 1997).

1.1.2 Leaf Level Variability

Variability in leaf level reflectance can be determined with Price's (1994) metrics D and θ, which compute the root mean square difference (D), and the angle difference (θ) between two spectra, allowing for a quantitative comparison between two spectra in multidimensional space. The use of these metrics demonstrates just how variable leaf reflectance can be, and not just between species. Spectral variation has been seen within the same leaf, leaves of the same branch and plant, and of different plants of the same species in close proximity (Williams 1991, Castro-Esau, Sánchez-Azofeifa et al. 2006), with even greater variability between the same species at different sites (Hesketh and Sánchez-Azofeifa 2012). This variability, along with interspecific similarity, confounds classification attempts based on reflectance spectra (Ollinger 2011). Despite this, an ever-increasing number of studies successfully use field or remotely sensed spectra for classification, with high accuracy. However, this accuracy does decline in sites with high diversity, such as the tropics: in fact there a linear relationship between accuracy and species number has been identified (Castro-Esau, Sánchez-Azofeifa et al. 2006).

Based on a study of 1449 species of Amazonian trees Asner, Martin et al. (2014) found that variability of spectral leaf reflectance between species is greatest in the near-infrared region (720 – 1400 nm) (NIR), followed by shortwave infrared (1400 – 2500 nm) (SWIR) to a lesser degree, though still showing strong variation. The visible wavelength region (400-700 nm) (VIS) demonstrated the lowest degree of variability, which has been attributed to the fact that pigments are abundant in healthy leaves and strongly absorb

in the visible region. Although the degree of variability seen amongst Asner's Amazonian samples is far greater than that shown by other studies, the trend of lower variability within the visible region in comparison to the longer wavelengths is consistent (Castro-Esau, Sánchez-Azofeifa et al. 2004).

Despite minimal variability, wavelength bands that fall within the visible region are frequently selected by various statistical techniques for discrimination of plant groups. In fact, many studies are solely reliant upon visible and near infrared regions as their spectral measurements do not extend beyond ~ 900 - 1000 nm. Bands selected from the visible region are often related to the blue-green edge, green peak, and red edge chlorophyll absorption features (Castro-Esau, Sánchez-Azofeifa et al. 2006). However, classification studies relying upon a selection of reflectance bands have utilised bands from across the optical spectrum, such as at wavelengths in the near-infrared and shortwave-infrared associated with fundamental bond vibrations (stretching and bending) in chemical constituents (Manley 2014). The range of bands selected throughout the 400 – 2500 nm wavelength range for different studies is influenced by the spectral data used and the taxa within the study.

1.1.3 Canopy Level Variability

Canopy level variability is influenced by the foliar reflectance properties previously mentioned, though it is dominated by the structure of the vegetation (Asner 1998, Gara, Darvishzadeh et al. 2018). Often this structure is described in a simplified manner as the Leaf Area Index (LAI), which is defined as the ratio between total single sided leaf area to ground area, and by Leaf Angle Distribution (LAD), which describes the gradient of leaf angles between horizontal and vertical that occur within the canopy.

LAI is a combination of a number of structural characteristics such as crown shape, canopy density and volume, clumping and gap fraction and has been successfully remotely sensed with normalised difference vegetation index (NDVI) products (Tesemma, Wei et al. 2014). These all affect the likelihood that a photon may be absorbed or scattered by the vegetation and the depth within the canopy that the interaction may occur. Increases in LAI have been shown to increase reflectance though limited to NIR region and saturated at high LAI values (Ollinger 2011). With the exception

of soil reflectance in sparse canopies LAI has the highest contribution to canopy reflectance variability (Asner 1998).

LAI is highly dependent on resource availability, showing high variation between poor and favourable growing conditions (Whitehead and Beadle 2004). Along with growing conditions, LAI can vary temporally, rapidly increasing with growth of young vegetation, before reaching a peak then lowering to a more constant value at maturity. The effect of seasonality is dependent on vegetation type with deciduous and annual species showing large degrees of cyclical variation, whilst evergreens show low to medium variability induced by seasonal variations in temperature, solar irradiance, and water availability (Tesemma, Wei et al. 2014). The relationship between LAI and environmental water conditions has been exploited to provide remotely sensed data for pre-dawn water potential, soil water potential, evapotranspiration, aridity index (Palmer, Fuentes et al. 2010), and soil conditions (Lausch, Zacharias et al. 2013).

The orientation of leaves within a canopy can have a dramatic effect on crown shape and therefore canopy reflectance. Estimated by LAD, this orientation has a large impact on the scattering and absorption of photons within the canopy. A canopy with predominantly vertically orientated leaves will demonstrate increased penetration of incident radiation resulting in increased scattering and absorption within the canopy, having the effect of lowering the canopy's overall reflectance. LAD has a greater effect on reflectance variability than LAI, with reflectance substantially decreasing as leaf angle increased from horizontal. Unlike LAI, LAD variability is not limited to NIR wavelengths, instead showing a change in reflectance across the spectrum, though to a lesser degree in VIS wavelengths (Ollinger 2011). Leaf angles vary depending on position within the canopy or latitude, with leaves in the upper canopy being more vertical than those in the mid and lower canopy, as well as leaves at the equator being the most horizontal (Miller and Lin 1985, Hollinger 1989, Huemmrich 2013).

1.1.4 Intra-specific Variability

In classification studies, it is common for spectral variation within taxa or functional groups to reduce the accuracy of classification or prevent the use of a model derived from a sampling site being used on another spatially or temporally distinct site. This

issue plays into the “one place, one time” problem identified by Woodcock (2002), where studies are often spatially and temporally constrained, limiting their insight and applicability to other related contexts. Varying edaphic and climatic conditions between sampling locations such as temperature, water availability, soil mineral nutrients, and solar irradiance are all known to influence geographic intra-specific variation (Carter 1993). Although this problem has been noted in many studies, few have attempted to quantify spectro-spatial variability and its relationship to the climatic and edaphic conditions that cause it.

Of studies that have, most have generally focused on changes along an elevation gradient. Although elevation is a single environmental variable in and of itself, it is correlated with many of the previously mentioned determinants of intra-specific variability. Due to this, studies such as Richardson, Berlyn et al. (2001), and Richardson and Berlyn (2002) relate observed spectro-spatial variability along an elevation gradient to stress, as the studied taxa approach the edges of their respective range limits. The phenotypic responses to the stressors included reduction in overall chlorophyll, as well as an increase in the carotenoid : chlorophyll a ratio, and a reduction in photosynthetic radiation efficiency. The Asner, Martin et al. (2014) study of Amazonian trees showed a similar trend, as productivity decreased with a decrease in photosynthetic pigments and foliar nitrogen along an elevation gradient. As well as elevation, the effect of soil fertility was examined amongst the Amazonian trees. It was found that trees from a low soil fertility environment contained low concentrations of macro- and micro-nutrients whilst having increased concentrations of defence compounds including phenols, tannins and lignin. This variability results in a significant decrease in classification accuracy, as demonstrated by Castro-Esau, Sánchez-Azofeifa et al. (2006), where classification within a site was successful, though classification between sites was not.

Changing of the seasons sees changes in precipitation, average temperature, and solar irradiance, resulting in phenological responses from plants such as leaf flush, flowering, fruiting, stress and senescence responses. As with the previously discussed spatial variability in intraspecific reflectance, temporal variability brought about by this suite of changes limits transferability of classifications to the season in which the spectra were recorded. This is highlighted by across season studies from tropical regions that have

recorded as much as a tenfold decrease in classification accuracy between wet and dry and seasons (Castro-Esau, Sánchez-Azofeifa et al. 2006, Hesketh and Sánchez-Azofeifa 2012).

However, contrary to this, some studies have utilised spectro-temporal variability to increase spectral separability of taxa, and therefore increase classification accuracy. For example, when studying the spectro-temporal variability caused by changes in the phenological cycle of English moorland plant functional groups, Cole, McMorrow et al. (2014) identified the optimal spectral indices to maximise spectral separation for each month. Amongst the moorland groups, the spring/summer months of April and July offered the greatest spectral separability between all groups using the Plant Senescence Reflection Index (PSRI). However, between specific groups other indices out-performed PSRI, such as Red Edge Position (REP) for the shrub and bryophytes group. As well as the optimal indices changing between groups, different indices optimally described variability for each month, with Photochemical Reflectance Index (PRI) and Cellulose Absorption Index (CAI) outperforming PSRI in June for all groups. The results of this study demonstrate the complex interactions between vegetation seasonal responses and spectral reflectance, and how knowledge of these responses amongst target samples can help to guide timing of spectral collection and wavelength selection. The Cole, McMorrow et al. (2014) study builds upon and expands the knowledge in this area, which so far has been limited to a small number of studies, such as for Australian seagrasses (Fyfe 2003), South African wetland trees (Van Deventer, Cho et al. 2013), C3 and C4 grasses (Wang, Hunt Jr et al. 2013), European boreal forests (Grigorieva, Brovkina et al. 2020), and Hawaiian invasive species (Asner, Martin et al. 2006).

A method to handle intra-class variability and alleviate the one place, one time problem would greatly increase the applicability of classification models, allowing them to utilise more classes from wider ranging spatial and temporal distances. Hyperspectral remote sensing is currently focused on improving classification results by including more spatial, structural and textural information to the analysis to aid in class differentiation. Despite this focus of the wider community, this thesis aims to improve the information that can be gained solely from spectral signals. For this reason, this thesis attempts to develop such a methodology, aided by significant advances in machine and deep learning. My

research investigates the use of data augmentation, the process of artificially expanding the sample size of a dataset as a way to handle intra-class variability and inter-class confusion amongst hyperspectral plant reflectance spectra. Data augmentation has become widespread in the machine learning community, particularly so for image classification as it offers novel views of a sample image to the training classification model. Whilst not as commonly used with 1-dimensional data its successful use for timeseries (Harada, Hayashi et al. 2019) sound synthesis (Donahue, McAuley et al. 2018), or anonymising medical data (Esteban, Hyland et al. 2017) demonstrates its potential.

1.2. Research Aims & Objectives

As outlined in section 1.1, hyperspectral data are capable of classifying vegetation to high levels of accuracy between closely related and functionally similar taxa, though is extremely susceptible to the relationship between intra and inter-class variability. I believe this is a significant roadblock to wider adoption of hyperspectral vegetation classification due to the increased data gathering and processing required to overcome it, as well as the associated reduced generalizability and transferability of classification models.

Therefore, it is the primary aim of this thesis to develop methods of improving the ratio between intra and inter-class variability in hyperspectral vegetation spectra, and ultimately increasing classification accuracy and reliability. An important criterion associated with this aim is to develop methods that do not require collection of new hyperspectral or supportive data so that they can be easily applied to previously collected hyperspectral data. The specific objectives of this thesis are threefold:

1. To assess the influence of class variability and methodologies of overcoming it by reviewing the last two decades of hyperspectral vegetation classification literature, particularly focusing on waveband selection frequency; waveband selection frequency variation by taxonomic, structural or functional group; and the influence of feature selection choice on resultant classification accuracy, generalizability, and transferability.
2. To determine the viability of synthesising hyperspectral plant data via a generative deep learning model for the purpose of data augmentation.

3. Refining the previous generative deep learning model, to produce transformed replications of hyperspectral plant classes with reduced intra-class variability whilst exaggerating inter-class differences.

1.3. Thesis Structure

This thesis comprises five chapters. Chapter one summarises the relevant literature, highlighting key knowledge gaps whilst providing context for the motivations behind this thesis and methodologies to be used. Drawing from this review, feature selection trials were performed to examine how study design (number of classes, number of samples, included species, and feature selection method) influenced waveband selection. Chapters two, three and four make up the body of research directly addressing the objectives of the thesis and are presented as published papers or manuscripts intended for publication. Chapter five consists of the discussion and conclusion, highlighting key results, presenting the significance and contributions of this thesis as well as suggesting future research opportunities. The following summaries highlight the content of each research chapter.

Chapter two: Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. Published as Hennessy, A., K. Clarke and M. Lewis (2020). "Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability." *Remote Sensing* 12(1): 113.

This chapter consists of a review of the hyperspectral vegetation classification literature from 1996 to 2018 focusing on the methods and results from the selection of informative spectral wavebands or regions. This study found large degrees of variability in the feature selection of spectral regions between studies, driven by study parameters beyond characteristics of the target classes, most specifically the choice of feature selection algorithm. These findings are based on both the review of the literature as well as feature selection experiments designed to determine which elements of the study design most contributed to variation seen in selected wavebands. It was found that this variability was enough to prevent determination of generalizable and therefore transferable spectral regions for specific taxonomic discrimination, though broad trends such as the importance of VIS and red edge wavelengths were apparent. As a result of

these findings caution was suggested in relying upon feature selection recommendations from the literature to guide waveband selections or classifications for new plant discrimination applications, as such recommendations appear to be weakly generalizable between studies.

Chapter three: Generative adversarial network synthesis of hyperspectral vegetation data. Published as Hennessy, A., K. Clarke and M. Lewis (2021). "Generative adversarial network synthesis of hyperspectral vegetation data." *Remote Sensing* 13(12): 2243.

This chapter aims to improve classification accuracy and generalisability by increasing training dataset sample sizes and improving class distributions to better reflect the evaluation dataset via data augmentation. This was achieved by applying advances in generative deep learning models to produce realistic synthetic hyperspectral vegetation data, whilst maintaining class relationships. Specifically, a generative adversarial network (GAN) was trained on two vegetation hyperspectral datasets, demonstrating the ability to approximate the distribution of the training samples. These synthetic samples are then combined with the original spectra to produce an augmented dataset for use in training supervised classification models. This chapter demonstrates that increasing the number of samples available for training classifiers can increase classification accuracy, though ultimately has minimal effect on the intra/inter-class ratio of the datasets.

Chapter four: Generative adversarial network-based style transfer for data augmentation of hyperspectral vegetation data. Submitted for publication as Hennessy, A., K. Clarke and M. Lewis (*Submitted*) “Generative adversarial network-based style transfer for data augmentation of hyperspectral vegetation data.” *Remote Sensing*.

This chapter aims to build upon the previous data augmentation and hyperspectral synthesis method to further reduce intra-class variability and to decrease the possibility of inter-class overlap in order to improve classification accuracy. This was achieved by adding the ability to transfer the style (reflectance characteristics) of a hyperspectral class to another class creating new transformed hybrid samples. These hybrid samples were synthesised as a replicate of each real sample with the GAN model, rather than synthesizing novel samples from a learned class distribution as previously done. Now with the capability of producing one-to-one representations of real samples the GAN

model can be used to introduce transformations into the replicated spectra, either by use of class specific GAN models or latent space interpolation. The data augmentation presented within this chapter directly addresses all three initial issues of low sample size, interspecies similarity and within-species variability, resulting in significant increases in classification accuracy.

Chapter five: Discussion and Conclusion

This chapter brings together and highlights the key findings of the previous chapters. It emphasises the significant contributions to the field of hyperspectral vegetation classification whilst additionally providing recommendations for future research.

1.4 References

Adam, E., O. Mutanga, D. Rugege and R. Ismail (2012). "Discriminating the papyrus vegetation (*Cyperus papyrus* L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP." *International Journal of Remote Sensing* 33(2): 552-569.

Alonso, M. C., J. A. Malpica and A. M. de Agirre (2011). Consequences of the Hughes phenomenon on some classification techniques. ASPRS 2011 Annual Conference, Milwaukee, Wisconsin May.

Alonzo, M., B. Bookhagen and D. A. Roberts (2014). "Urban tree species mapping using hyperspectral and lidar data fusion." *Remote Sensing of Environment* 148: 70-83.

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

Asner, G. P., R. E. Martin, K. M. Carlson, U. Rascher and P. M. Vitousek (2006). "Vegetation–climate interactions among native and invasive species in Hawaiian rainforest." *Ecosystems* 9(7): 1106-1117.

Asner, G. P., R. E. Martin, R. Tupayachi, C. B. Anderson, F. Sinca, L. Carranza-Jiménez and P. Martinez (2014). "Amazonian functional diversity from forest canopy chemical assembly." *Proceedings of the National Academy of Sciences* 111(15): 5604-5609.

Bolch, E. A., M. J. Santos, C. Ade, S. Khanna, N. T. Basinger, M. O. Reader and E. L. Hestir (2020). Remote Detection of Invasive Alien Species. *Remote Sensing of Plant Biodiversity*. J. Cavender-Bares, J. A. Gamon and P. A. Townsend. Cham, Springer International Publishing**:** 267-307.

Carter, G. A. (1993). "Responses of leaf spectral reflectance to plant stress." *American Journal of Botany* 80(3): 239-243.

Castro-Esau, K. L., G. Sánchez-Azofeifa and T. Caelli (2004). "Discrimination of lianas and trees with leaf-level hyperspectral data." *Remote Sensing of Environment* 90(3): 353-372.

Castro-Esau, K. L., G. A. Sánchez-Azofeifa, B. Rivard, S. J. Wright and M. Quesada (2006). "Variability in leaf optical properties of Mesoamerican trees and the potential for species classification." *American Journal of Botany* 93(4): 517-530.

Cavender-Bares, J., J. E. Meireles, J. J. Couture, M. A. Kaproth, C. C. Kingdon, A. Singh, S. P. Serbin, A. Center, E. Zuniga and G. Pilz (2016). "Associations of leaf spectra with genetic and phylogenetic variation in oaks: prospects for remote detection of biodiversity." *Remote Sensing* 8(3): 221.

Cho, M. A., P. Debba, R. Mathieu, L. Naidoo, J. Van Aardt and G. P. Asner (2010). "Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: Canopy-level analysis." *IEEE Transactions on Geoscience and Remote Sensing* 48(11): 4133-4142.

Clark, M. L., D. A. Roberts and D. B. Clark (2005). "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales." *Remote Sensing of Environment* 96(3): 375-398.

Cole, B., J. McMorrow and M. Evans (2014). "Spectral monitoring of moorland plant phenology to identify a temporal window for hyperspectral remote sensing of peatland." *ISPRS Journal of Photogrammetry and Remote Sensing* 90: 49-58.

Curran, P. J. (1989). "Remote sensing of foliar chemistry." *Remote Sensing of Environment* 30(3): 271-278.

Conel, J.E., R.O. Green, G. Vane, C.J. Bruegge, R.E. Alley (1987) “AIS-2 radiometry and a comparison of methods for the recovery of ground reflectance.” In Proceedings of the 3rd Airborne Imaging Data Analysis Workshop, Pasadena, CA, USA, 2-4 June 1987; pp. 18-47.

Dalponte, M., L. Bruzzone and D. Gianelle (2012). "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data." *Remote Sensing of Environment* 123: 258-270.

Dalponte, M., L. Bruzzone, L. Vescovo and D. Gianelle (2009). "The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas." *Remote Sensing of Environment* 113(11): 2345-2355.

Datt, B. (2000). Recognition of eucalyptus forest species using hyperspectral reflectance data. Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International, IEEE.

Deng, W., Y. Huang, C. Zhao, L. Chen and X. Wang (2016). "Bayesian discriminant analysis of plant leaf hyperspectral reflectance for identification of weeds from cabbages." *African Journal of Agricultural Research* 11(7): 551-562.

Donahue, C., J. McAuley and M. Puckette (2018). "Adversarial audio synthesis." *arXiv preprint arXiv:1802.04208*.

Esteban, C., S. L. Hyland and G. Rätsch (2017). "Real-valued (medical) time series generation with recurrent conditional gans." *arXiv preprint arXiv:1706.02633*.

Fassnacht, F. E., H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, C. Straub and A. Ghosh (2016). "Review of studies on tree species classification from remotely sensed data." *Remote Sensing of Environment* 186: 64-87.

Fassnacht, F. E., C. Neumann, M. Förster, H. Buddenbaum, A. Ghosh, A. Clasen, P. K. Joshi and B. Koch (2014). "Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6): 2547-2561.

Fernandes, A., P. Melo-Pinto, B. Millan, J. Tardaguila and M. Diago (2015). "Automatic discrimination of grapevine (Vitis vinifera L.) clones using leaf hyperspectral imaging and partial least squares." *The Journal of Agricultural Science* 153(3): 455-465.

Ferreira, M. P., A. E. B. Grondona, S. B. A. Rolim and Y. E. Shimabukuro (2013). "Analyzing the spectral variability of tropical tree species using hyperspectral feature selection and leaf optical modeling." *Journal of Applied Remote Sensing* 7(1): 073502-073502.

Ferreira, M. P., M. Zortea, D. C. Zanotta, Y. E. Shimabukuro and C. R. de Souza Filho (2016). "Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data." *Remote Sensing of Environment* 179: 66-78.

Fyfe, S. (2003). "Spatial and temporal variation in spectral reflectance: Are seagrass species spectrally distinct?" *Limnology and Oceanography* 48(1part2): 464-479.

Gao, B.-C., K. B. Heidebrecht and A. F. Goetz (1993). "Derivation of scaled surface reflectances from AVIRIS data." *Remote sensing of Environment* 44(2-3): 165-178.

Gara, T. W., R. Darvishzadeh, A. K. Skidmore and T. Wang (2018). "Impact of Vertical Canopy Position on Leaf Spectral Properties and Traits across Multiple Species." *Remote Sensing* 10(2): 346.

Goetz, A. F., J. W. Boardman, B. C. Kindel and K. B. Heidebrecht (1997). Atmospheric corrections: On deriving surface reflectance from hyperspectral imagers. SPIE Proceedings, Vol. 3118 (1997), pp. 14-22

Goodwin, N., R. Turner and R. Merton (2005). "Classifying Eucalyptus forests with high spatial and spectral resolution imagery: an investigation of individual species and vegetation communities." *Australian Journal of Botany* 53(4): 337-345.

Grigorieva, O., O. Brovkina and A. Saidov (2020). “An original method for tree species classification using multitemporal multispectral and hyperspectral satellite data.” *Silva Fennica.* 54. 10.14214/sf.10143

Guo, X., R. Wang, J. Cheng and Q. Wang (1998). Radiometric restoration of airborne imaging spectrometer data. Hyperspectral Remote Sensing and Application, International Society for Optics and Photonics.

Harada, S., H. Hayashi and S. Uchida (2019). "Biosignal Generation and Latent Variable Analysis With Recurrent Generative Adversarial Networks." *IEEE Access* : 1-1.

Held, A., S. Phinn, M. Soto-Berelov and S. Jones (2015). "AusCover good practice guidelines: A technical handbook supporting calibration and validation activities of remotely sensed data products." *TERN AusCover* 352.

Hesketh, M. and G. A. Sánchez-Azofeifa (2012). "The effect of seasonal spectral variation on species classification in the Panamanian tropical forest." *Remote Sensing of Environment* 118: 73-82.

Hollinger, D. (1989). "Canopy organization and foliage photosynthetic capacity in a broad-leaved evergreen montane forest." *Functional Ecology 3*: 53-62.

Huemmrich, K. F. (2013). "Simulations of seasonal and latitudinal variations in leaf inclination angle distribution: Implications for remote sensing." Advances in Remote Sensing 2: 93-101

Hueni, A. (2006). Field spectroradiometer data: acquisition, organisation, processing and analysis on the example of New Zealand native plants: a thesis presented in fulfilment of the requirements for the degree of Master of Philosophy in Earth Science, Massey University, Palmerston North, New Zealand.

Hughes, G. (1968). "On the mean accuracy of statistical pattern recognizers." *IEEE Transactions on Information Theory* 14(1): 55-63.

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

Jones, T. G., N. C. Coops and T. Sharma (2010). "Employing ground-based spectroscopy for tree-species differentiation in the Gulf Islands National Park Reserve." *International Journal of Remote Sensing* 31(4): 1121-1127.

Knipling, E. B. (1970). "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation." *Remote Sensing of Environment* 1(3): 155-159.

Kruse, F. A. (1988). "Use of airborne imaging spectrometer data to map minerals associated with hydrothermally altered rocks in the northern grapevine mountains, Nevada, and California." *Remote Sensing of Environment* 24(1): 31-51.

Lausch, A., S. Zacharias, C. Dierke, M. Pause, I. Kühn, D. Doktor, P. Dietrich and U. Werban (2013). "Analysis of vegetation and soil patterns using hyperspectral remote sensing, EMI, and gamma-ray measurements." *Vadose Zone Journal* 12(4): 1-15.

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

Lewis, M. (2002). "Spectral characterization of Australian arid zone plants." *Canadian Journal of Remote Sensing* 28(2): 219-230.

Manley, M. (2014). "Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials." *Chemical Society Reviews* 43(24): 8200-8214.

Mariotto, I., P. S. Thenkabail, A. Huete, E. T. Slonecker and A. Platonov (2013). "Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HyspIRI mission." *Remote Sensing of Environment* 139: 291-305.

Miller, D. and J. Lin (1985). Canopy architecture of a red maple edge stand measured by a point drop method. The forest-atmosphere interaction, Springer**:** 59-70.

Ollinger, S. V. (2011). "Sources of variability in canopy reflectance and the convergent properties of plants." *New Phytol* 189(2): 375-394.

Pal, M. and G. M. Foody (2010). "Feature selection for classification of hyperspectral data by SVM." *IEEE Transactions on Geoscience and Remote Sensing* 48(5): 2297-2307.

Palmer, A. R., S. Fuentes, D. Taylor, C. Macinnis-Ng, M. Zeppel, I. Yunusa and D. Eamus (2010). "Towards a spatial understanding of water use of several land-cover classes: an examination of relationships amongst pre-dawn leaf water potential, vegetation water use, aridity and MODIS LAI." *Ecohydrology: Ecosystems, Land and Water Process Interactions Ecohydrogeomorphology* 3(1): 1-10.

Papp, L., B. van Leeuwen, P. Szilassi, Z. Tobak, J. Szatmári, M. Árvai, J. Mészáros and L. Pásztor (2021). "Monitoring Invasive Plant Species Using Hyperspectral Remote Sensing Data." *Land* 10(1): 29.

Peerbhay, K. Y., O. Mutanga and R. Ismail (2013). "Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa." *ISPRS Journal of Photogrammetry and Remote Sensing* 79: 19-28.

Price, J. C. (1994). "How unique are spectral signatures?" *Remote Sensing of Environment* 49(3): 181-186.

Prospere, K., K. McLaren and B. Wilson (2014). "Plant species discrimination in a tropical wetland using in situ hyperspectral data." *Remote Sensing* 6(9): 8494-8523.

Pu, R. (2009). "Broadleaf species recognition with in situ hyperspectral data." *International Journal of Remote Sensing* 30(11): 2759-2779.

Richards, J.A. (1999) “Remote Sensing Digital Image Analysis: An Introduction”, Springer-Verlag, Berlin, Germany, 240.

Richardson, A. D. and G. P. Berlyn (2002). "Spectral reflectance and photosynthetic properties of Betula papyrifera (Betulaceae) leaves along an elevational gradient on Mt. Mansfield, Vermont, USA." *American Journal of Botany* 89(1): 88-94.

Richardson, A. D., G. P. Berlyn and T. G. Gregoire (2001). "Spectral reflectance of Picea rubens (Pinaceae) and Abies balsamea (Pinaceae) needles along an elevational gradient, Mt. Moosilauke, New Hampshire, USA." *American Journal of Botany* 88(4): 667-676.

Richter, R., B. Reu, C. Wirth, D. Doktor and M. Vohland (2016). "The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area." *International Journal of Applied Earth Observation and Geoinformation* 52: 464-474.

Richter, R., D. Schläpfer and A. Müller (2011). "Operational Atmospheric Correction for Imaging Spectrometers Accounting for the Smile Effect." *Geoscience and Remote Sensing, IEEE Transactions on* 49: 1772-1780.

Roberts, D., Y. Yamaguchi and R. Lyon (1986). "Comparison of various techniques for calibration of AIS data." *NASA STI/Recon Technical Report N* 87: 21-30.

Salisbury, J. W. (1998). Spectral measurements field guide, Earth Satellite Corp Checy Chade MD.

Schläpfer, D., A. Hueni and R. Richter (2018). "Cast Shadow Detection to Quantify the Aerosol Optical Thickness for Atmospheric Correction of High Spatial Resolution Optical Imagery." *Remote Sensing* 10(2): 200.

Shang, X. and L. A. Chisholm (2014). "Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6): 2481-2489.

Sommer, C., S. Holzwarth, U. Heiden, M. Heurich, J. Müller and W. Mauser (2015). Feature based tree species classification using hyperspectral and lidar data in the Bavarian Forest National Park. EARSel eProceedings, Vol. 14, Special Issue 2: 9th EARSeL Imaging Spectroscopy Workshop, 2015.

Tesemma, Z. K., Y. Wei, A. W. Western and M. C. Peel (2014). "Leaf area index variation for crop, pasture, and tree in response to climatic variation in the Goulburn–Broken catchment, Australia." *Journal of Hydrometeorology* 15(4): 1592-1606.

Thenkabail, P. S., E. A. Enclona, M. S. Ashton and B. Van Der Meer (2004). "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications." *Remote Sensing of Environment* 91(3): 354-376.

Thenkabail, P. S., I. Mariotto, M. K. Gumma, E. M. Middleton, D. R. Landis and K. F. Huemmrich (2013). "Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/EO-1 data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6(2): 427-439.

Thompson, D. R., V. Natraj, R. O. Green, M. C. Helmlinger, B.-C. Gao and M. L. Eastwood (2018). "Optimal estimation for imaging spectrometer atmospheric correction." *Remote Sensing of Environment* 216: 355-373.

Tirado, S. B., S. St Dennis, T. A. Enders and N. M. Springer (2020). "Utilizing top-down hyperspectral imaging for monitoring genotype and growth conditions in maize." *bioRxiv*: 2020.2001.2021.914069.

Underwood, E., S. Ustin and D. DiPietro (2003). "Mapping nonnative plants using hyperspectral imagery." *Remote Sensing of Environment* 86(2): 150-161.

Van Deventer, H., M. A. Cho and O. Mutanga (2013). "Do seasonal profiles of foliar pigments improve species discrimination of evergreen coastal tree species in KwaZulu-Natal, South Africa?" Proc Confer. 35th Internation Symposium Remote Sensing Environment (ISRSE), pp. 1-12

Wang, C., E. R. Hunt Jr, L. Zhang and H. Guo (2013). "Phenology-assisted classification of C3 and C4 grasses in the US Great Plains and their climate dependency with MODIS time series." *Remote Sensing of Environment* 138: 90-101.

Wang, J., R. Xu and S. Yang (2009). "Estimation of plant water content by spectral absorption features centered at 1,450 nm and 1,940 nm regions." *Environmental Monitoring and Assessment* 157(1-4): 459-469.

Whitehead, D. and C. L. Beadle (2004). "Physiological regulation of productivity and water use in Eucalyptus: a review." *Forest Ecology and Management* 193(1-2): 113-140.

Williams, D. L. (1991). "A comparison of spectral reflectance properties at the needle, branch, and canopy level for selected conifer species." *Remote Sensing of Environment* 35(2-3): 79-93.

Woodcock, C. E. (2002). "Uncertainty in remote sensing." *Uncertainty in Remote Sensing and GIS*: John Wiley & Sons, 19-24.

1

CHAPTER TWO: Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability

This chapter is published as:

Hennessy, A., K. Clarke and M. Lewis (2020). "Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability." *Remote Sensing* 12(1): 113.

This chapter appears as published with minor modifications to reference style and formatting of figure and table captions for consistency with the remainder of the thesis

Statement of Authorship

Title of Paper	Hyperspectral classification of plants: a review of waveband selection generalisability	
Publication Status	☒ Published	☐ Accepted for Publication
	☐ Submitted for Publication	☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Hennessy, A.; Clarke, K.; Lewis, M. Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability. Remote Sens. 2020, 12, 113, https://doi.org/10.3390/rs12010113.	

Principal Author

Name of Principal Author (Candidate)	Andrew Hennessy	
Contribution to the paper	Conceptualization, Methodology, data curation, formal analysis, writing – original draft preparation, writing – review and editing.	
Overall percentage (%)	70%	
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature		Date: 13/07/2021

Co-author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kenneth D. Clarke	
Contribution to the paper	Methodology, formal analysis, writing – review and editing	
Signature		Date: 13/07/2021

Name of Co-Author	Megan M. Lewis	
Contribution to the paper	Methodology, formal analysis, writing – review and editing	
Signature		Date: 12/07/2021

Abstract

Hyperspectral sensing, measuring reflectance over visible to shortwave infrared wavelengths, has enabled the classification and mapping of vegetation at a range of taxonomic scales, often down to species level. Classification with hyperspectral measurements, acquired by narrow band spectroradiometers or imaging sensors, has generally required some form of spectral feature selection to reduce the dimensionality of the data to a level suitable for the construction of a classification model. Despite the large number of hyperspectral plant classification studies, an in-depth review of feature selection methods and resultant waveband selections has not yet been performed. Here we present a review of the last 22 years of hyperspectral vegetation classification literature that evaluates overall waveband selection frequency; waveband selection frequency variation by taxonomic, structural or functional group; and the influence of feature selection choice by comparing methods such as Stepwise Discriminant Analysis (SDA), Support Vector Machines (SVM), and Random Forests (RF).

This review determined that all characteristics of hyperspectral plant studies influence the wavebands selected for classification. This includes the taxonomic, structural and functional groups of the target samples, the methods and scale at which hyperspectral measurements are recorded, and the feature selection method used. Furthermore, these influences do not appear to be consistent. Moreover, the considerable variability in waveband selection caused by the feature selectors effectively masks analysis of any variability between studies related to plant groupings. Additionally, questions are raised about the suitability of SDA as a feature selection method, with it producing waveband selections at odds with the other feature selectors.

Caution is recommended when choosing a feature selector for hyperspectral plant classification: we recommend multiple methods being performed. The resultant sets of selected spectral features can be either evaluated individually by multiple classification models or combined as an ensemble for evaluation by a single classifier. Additionally, we suggest caution in relying upon waveband recommendations from the literature to guide waveband selections or classifications for new plant discrimination applications, as such recommendations appear to be weakly generalizable between studies.

2.1. Introduction

Classification of reflectance spectra to determine broad plant type or species has been explored increasingly over the past two decades. This has been driven by the increased availability of hyperspectral sensing from imaging spectrometers and field spectroradiometers, and increasing need from environmental conservation, agriculture and forestry groups (Fassnacht, Latifi et al. 2016). High classification accuracies, particularly at fine taxonomic units such as species or even clones for grapevine varieties (Fernandes, Melo-Pinto et al. 2015) has in some cases been enabled by hyperspectral observation (Clark, Roberts et al. 2005) Hyperspectral measurements have been used to classify a variety of plant types including annual gramineous weeds (Deng, Huang et al. 2016), food crops (Mariotto, Thenkabail et al. 2013), arid zone shrubs (Lewis 2002) and montane/sub-alpine trees (Sommer, Holzwarth et al. 2015), growing in equally varied environments from tropical wetlands (Prospere, McLaren et al. 2014), urban streetscapes (Alonzo, Bookhagen et al. 2014), savanna plains (Cho, Debba et al. 2010), and alpine forests (Dalponte, Bruzzone et al. 2012). Due to the scale required to map and monitor the world's vegetation, fast, generalizable, and objective methods that provide results that can be quickly and easily shared and analysed are required. Hyperspectral imagery and data can fulfil these requirements, producing digital measurements that can be easily shared and quickly analysed with semi-automated procedures in a repeatable and objective manner. However, the potential generalisability of classification models has yet to be fully evaluated.

Hyperspectral measurements consist of numerous, finely spaced, contiguous measurements (wavebands) providing considerably more information about targets than broadband multispectral observations. These advantages come at the cost of high dimensionality and large data volumes. Hyperspectral instruments record radiance within the range 350 to 2500 nm of the electromagnetic spectrum, with bandwidths often between 1 and 10 nm. The number of wavebands per observation varies from hundreds to thousands. Training a classification model with such large numbers of spectral features requires a large sample size. However, since the collection of samples for hyperspectral studies is onerous, with high costs for imagery and arduous fieldwork for gathering field measurements, sample sizes tend to be small. Data of this high

dimensionality is prone to the Hughes phenomenon, also known as the curse of dimensionality, where an increasing number of features originally aids in improving classification, before the addition of more features decreases performance as noise and sparsity of the feature space increases (Alonso, Malpica et al. 2011). This problem is exacerbated by small sample sizes (Hughes 1968).

In order to overcome this, either the dimensionality of the data must first be reduced or a classification method less susceptible to the curse of dimensionality used (Pal and Foody 2010). Hyperspectral measurements tend to include noisy or redundant features, with high levels of collinearity between wavebands. The elimination of collinearity can substantially improve classification efforts and is in fact a requirement of parametric statistical methods that assume independence of all variables (Fassnacht, Neumann et al. 2014, Richter, Reu et al. 2016). Additionally, feature selection inherently reveals the spectral regions that offer the greatest discriminatory power for a set of samples. Long held associations between specific spectral regions or individual wavebands and biophysical or biochemical foliar traits (Curran 1989) have often guided researchers in selecting features to differentiate species or plant types. The overall aim of this review is to assess these assumptions in light of evidence from 22 years of hyperspectral plant studies.

2.1.1. Review Scope and Approach

Here we address some important questions that motivate much hyperspectral plant research. Do the taxonomic, structural or functional characteristics of plant types or species influence the spectral regions that are most important in classification, or are particular spectral regions consistently selected across a diversity of plant or ecological types? Review of selected features from the hyperspectral literature could identify best practices for feature selection methods, as well as detect wave-regions of high-utility, those that best generalize across taxonomic or ecological boundaries.

The search for literature spanned two decades from January 1996 to December 2018, focusing on peer reviewed journals in the English language. Search was performed with Google Scholar using combinations of the keywords *Hyperspectral, Spectra, Vegetation, Plant, Tree, Species, Identi*, Discriminat*, Classif*, Map, Feature Select*, Waveband,*

Band, UAV, Drone. In order to be included a study must have performed a feature selection technique on hyperspectral vegetation data with an aim to classify plant samples.

Many studies fulfilled the initial requirement, but did not report selected wavebands with sufficient specificity, and therefore could not be included. Here we present waveband selections derived from 38 hyperspectral vegetation classification studies. When applicable, studies that included multiple feature selection techniques were broken into sub-studies, increasing the total number of reviewed studies to 61 (Table 1, 2). These included studies are from a wide variety of scales (leaf, branch, canopy), recording methods (lab, field, aerial, satellite), taxonomic units, and bandwidths.

Additionally, a dataset was synthesised from hyperspectral measurements of 22 species of New Zealand plants collected as field spectra from 4 locations on the north island (Hueni 2006). This dataset was used to examine how study design (number of classes, number of samples, included species, and feature selection method) influenced waveband selection. This was performed with the aim of determining which elements of the study design most contributed to variation seen in selected wavebands.

The remainder of this paper is structured in the following way: Section 2 provides a meta-analysis of the selected wavebands, broken down by spectral region; Section 3 identifies and describes feature selection techniques from these studies, and where possible highlights their effects on waveband selection; Section 4 examines study design influence on waveband selection; Section 5 presents a synthesis of results and conclusions.

Table 1. Overview of VIS/NIR studies included in this review.

References	Wavelengths/Bandwidths	Classes	Pre-processing	Feature Selection Method	No. bands selected	Accuracy %	Study Context and Spatial Scale or Resolution
Aneece and Epstein	350 – 1025 nm, 3 nm	12	Band depth	Segmented PCA	12	77.0	Successional plant communities from canopy field spectra
Cao et al. (2018)	454 – 950 nm, 4 nm	8	Smoothing	SDA	14	91.4	Mangrove forest field canopy spectra
Cao et al. (2018)	454 – 950 nm, 4 nm	8	Smoothing	CFS	23	92.3	Mangrove forest field canopy spectra
Cao et al. (2018)	454 – 950 nm, 4 nm	8	Smoothing	SPA	23	93.1	Mangrove forest field canopy spectra
Cho et al. (2010)	384.8 – 1054.3 nm, 9.23	10		SAM Band Selector Addon	31	53.0	Savanna tree species from airborne imagery (1.12 m)
Dalponte et al. (2012)	403 – 989 nm, 4.6 nm	8		Sequential Forward Floating Selection	43	74.1	Alpine tree species and 2 non-species classes, airborne imagery (1 m)
Dian et al. (2014)	400 – 900 nm, 1 nm	13		Spec angle and dist., feature parameters	7	96.2	Varied plant species from lab leaf spectra
Eddy et al. (2014)	400 – 1000 nm, 10 nm	5		PCA, SDA, Manual selection	7	~91.4	Crop and weed species from field imagery (1.25 m)
Fung et al. (2003)	400 – 900 nm, 2.6 nm	25	Smoothing	Hierarchical Clustering	13	89.0	Sub-tropical tree species from lab leaf spectra
Gross and Heumann	475 – 900 nm, 1 nm	22		Forward Feature Selection	8	43.0	Herbaceous wetland species from field leaf spectra
Hoa et al. (2017)	325 – 1075 nm, 2 nm	6	Smoothing	SDA	6	92.0	Mangrove forest field canopy spectra
Hoa et al. (2017)	325 – 1075 nm, 2 nm	6	Smoothing, CR	SDA	17	93.6	Mangrove forest field canopy spectra
Lewis (2002)	400 – 900 nm, 1.4 nm	8		PCA, Discriminant Analysis	13	57.0	Arid zone plant groups from field leaf spectra
Naidoo et al. (2012)	384.8 – 1054.3, 9.23 nm	9		Random Forest, Gini Index	8	80.3	Savanna tree species from airborne imagery (1.3 m)
Naidoo et al. (2012)	384.8 – 1054.3, 9.23 nm	9	Continuum removed	Random Forest, Gini Index	9	~79.0	Savanna tree species from airborne imagery (1.3 m)
Peerbhay et al. (2013)	393 – 900 nm, 2.2 nm	6		PLSDA VIP score	78	88.8	Forestry species from airborne imagery (2.4 m)
Pu et al. (2012)	400 – 800 nm, 3 nm	3		Two Sample T-test	5	69.1	Seagrass species field canopy spectra
Pu et al. (2012)	400 – 800 nm, 3 nm	3	Normalized	Two Sample T-test	5	66.0	Seagrass species field canopy spectra
Pu et al. (2012)	400 – 800 nm, 3 nm	3	Normalized 1st Derivative	Two Sample T-test	5	71.1	Seagrass species field canopy spectra
Pu et al. (2012)	400 – 800 nm, 3 nm	3	Normalized 2nd Derivative	Two Sample T-test	5	73.2	Seagrass species field canopy spectra
Pu et al. (2012)	400 – 800 nm, 3 nm	3	1st Derivative	Two Sample T-test	5	69.1	Seagrass species field canopy spectra
Pu et al. (2012)	400 – 800 nm, 3 nm	3	2nd Derivative	Two Sample T-test	5	67.0	Seagrass species field canopy spectra
Sommer et al. (2015)	400 – 1000 nm, 3 nm	13	Normalisation	PCA, Correlation matrix, Band variance	53	77.0	European forest trees species from airborne imagery (1.6 m)

Table 2. Overview of VIS/SWIR studies included in this review.

References	Wavelengths/Bandwidth	Class	Pre-processing	Feature Selection Method	Bands	Accuracy %	Study Context and Spatial Scale or Resolution
Adam and Mutanga	350 – 2500 nm @ 3, 10	4		ANOVA, CART	8	97.4	Wetland species from field canopy spectra
Adam et al. (2012)	350 – 2500 nm @ 3, 10	4	Resampled	Random Forest	10	90.5	Wetland species from field canopy spectra
Alonzo et al. (2014)	385 – 2450 nm @ 9.6 nm	29		Forward Feature Selection	7	79.2	Urban street tree species from airborne imagery (3.7 m)
Aneece et al. (2018)	427 – 2355 nm @ 10 nm	4		PCA	15	86.3	Agricultural crops, Hyperion (30 m)
Beh et al. (2017)	350 – 2500 nm @ 10nm	6	Resampled	ANOVA, LDA	26	77.0	Mangrove species leaf scale
Chan and Paelinckx	400 – 2500 nm @ 16 nm	16		Best-First Search Algorithm	21	~69.5	Temperate forest ecotopes from airborne imagery (4 m)
Chan and Paelinckx	400 – 2500 nm @ 16 nm	16		Random Forest	21	~69.5	Temperate forest ecotopes from airborne imagery (4 m)
Das et al. (2018)	350 – 2350 nm @ 1 nm	14		ANOVA (Tukey HSD), CART	17	98.0	Rice genotypes from canopy spectra
Datt (2000)	400 – 2500 nm @ 10 nm	7	Resampled	Stepwise Discriminant Analysis	12	70.4	Eucalypt forest species from lab leaf spectra
Datt (2000)	400 – 2500 nm @ 10 nm	7	Resampled, 1st Derivative	Stepwise Discriminant Analysis	13	72.4	Eucalypt forest species from lab leaf spectra
Deng et al. (2016)	350 – 2500 nm @ 1.4, 2	7		PCA	8	84.3	Cabbage crops and weed species from field canopy spectra
Fernandes et al. (2013)	350 – 2450 nm @ 3,	4		Kruskal-Wallis post hoc Dunn, CART	56	~95	Giant Reed and coexisting vegetation from field canopy spectra
Ferreira et al. (2016)	400 – 2400 nm @ 4, 6	8	Smoothing	Stepwise Regression Wrapper	30	~70.0	Tropical tree species from airborne imagery (1 m)
George et al. (2014)	400 – 2350 nm @ 10 nm	6	Continuum Removed	Stepwise Discriminant Analysis	29	82.3	Himalayan forest species from satellite imagery (30 m)
Jones et al. (2010)	429 – 2400 nm @ 2 nm	11		Stepwise Discriminant Analysis	40	~98.0	Canadian forest tree species from lab leaf spectra
Jones et al. (2010)	429 – 2400 nm @ 2 nm	11	1st Derivative	Stepwise Discriminant Analysis	40	~98.0	Canadian forest tree species from lab leaf spectra
Jones et al. (2010)	429 – 2400 nm @ 2 nm	11	2nd Derivative	Stepwise Discriminant Analysis	40	~98.0	Canadian forest tree species from lab leaf spectra
Mariotto et al. (2013)	426.5 – 2355 nm @ 10	5		LS-means, SDA, PCA, LL-R^2	29	90.2	Crop species from satellite imagery (30 m)
Mariotto et al. (2013)	426.5 – 2355 nm @ 10	5	Resampled	LS-means, SDA, PCA, LL-R^2	21	92.0	Crop species from canopy field spectra
Papes et al. (2010)	415 – 2340 nm @ 10 nm	5		Stepwise Discriminant Analysis	25	100	Amazon tree species from satellite imagery (30 m)
Papes et al. (2010)	415 – 2340 nm @ 10 nm	5		Stepwise Discriminant Analysis	25	100	Amazon tree species from satellite imagery (30 m)
Prospere et al. (2014)	400 – 2400 nm @ 5 nm	46	Smoothing, Normalization	PCA	20	82.6	Tropical wetland species from field leaf spectra
Prospere et al. (2014)	400 – 2400 nm @ 5 nm	46	Smoothing, Normalization	Mann-Whitney *U*-test	21	86.8	Tropical wetland species from field leaf spectra
Prospere et al. (2014)	400 – 2400 nm @ 5 nm	46	Smoothing, Normalization	ANOVA	23	83.4	Tropical wetland species from field leaf spectra
Prospere et al. (2014)	400 – 2400 nm @ 5 nm	46	Smoothing, Normalization	SVM	20	87.1	Tropical wetland species from field leaf spectra
Prospere et al. (2014)	400 – 2400 nm @ 5 nm	46	Smoothing, Normalization	Random Forest	20	86.1	Tropical wetland species from field leaf spectra
Prospere et al. (2014)	400 – 2400 nm @ 5 nm	46	Smoothing, Normalization	Random Forest (a)	20	84.8	Tropical wetland species from field leaf spectra
Raczko et al. (2018)	413 – 2440 nm @ 0.6, 11	6		PCA	40	87.0	
Schmidt and Skidmore	400 – 2500 nm @ 2, 6,	27	Smoothing, Continuum removed	Mann-Whitney *U*-test, Manual Selection	6	-	Saltmarsh vegetation types from field canopy spectra
Shang and Chisholm	350 – 2500 nm @ 3, 10	7		ANOVA – post hoc Tukey-Kramer	9	94.7	Australian forest species from lab leaf spectra

Thenkabail et al. (2004)	390 – 2360 nm @ 10 nm	4	Resampled	PCA, LL-R^2, SDA, DGVI	22	97.0	Crops and savanna cover types from field canopy spectra
Thenkabail et al. (2013)	350 – 2350 nm @ 10 nm	8		Stepwise Discriminant Analysis	20	95.0	Crop types from field canopy spectra
Vaiphasa et al. (2005)	350 – 2500 nm @ 3, 10	16		Genetic Algorithm	4	~80.0	Mangrove species from lab leaf spectra
Vaiphasa et al. (2007)	350 – 2500 nm @ 3, 10	16		Genetic Algorithm	(30*4)	~80.0	Mangrove species from lab leaf spectra
Van Aardt and Wynne	400 – 2500 nm @ 10 nm	3		Stepwise Discriminant Analysis	10	65.0	Pine tree species from airborne imagery (3.4 m)
Van Aardt and Wynne	400 – 2500 nm @ 10 nm	3	1^{st} Derivative	Stepwise Discriminant Analysis	10	77.0	Pine tree species from airborne imagery (3.4 m)
Van Aardt and Wynne	400 – 2500 nm @ 10 nm	3	2^{nd} Derivative	Stepwise Discriminant Analysis	10	72.	Pine tree species from airborne imagery (3.4 m)
Wang et al. (2009)	350 – 2500 nm @ 10 nm	3	Resampled	ANOVA, Linear Discriminant Analysis	15	90.0	Mangrove species from lab leaf spectra

2.2. Meta-Analysis

The delineation of spectral regions in this review follows that of (Clark, Roberts et al. 2005) as adapted from (Asner 1998). Imaging and non-imaging hyperspectral instruments have different sampling intervals, so direct comparison of selected wavebands between studies is not possible. This was resolved by aggregating selected wavebands into 50 nm bins based on their band centres (Figure 1). The design and presentation of the binned wavelengths is adapted from (Fassnacht, Latifi et al. 2016) with adjustments. Additionally, the bin size of the histogram has the benefit of grouping highly correlated and often redundant wavebands together, reducing noise from selection of correlated features from the analysis. The percentage of studies that selected wavebands within each 50 nm region is presented in the histogram, giving the selection rate for each 50 nm spectral region. The binned table, and selection rate histogram (Figure 1) only gives an indication of the rate with which a spectral region was selected and does not include information on the number of bands selected in each 50 nm bin nor the determined importance of a selected band for subject discrimination.

2.2.1. Spectral Range

Of the studies that met the rules for inclusion in this review, 38 used hyperspectral data spanning most of the range 350-2500 nm. However, a number of studies utilised devices that recorded a more restricted wavelength range between 350 and 1100 nm, generally from 400 nm to 800 nm or 1000 nm (Table 1). These studies are presented separately as the absence of SWIR and much of the NIR has shown to have an influence on waveband selection for the VIS and partial NIR (Rivard, Sanchez-Azofeifa et al. 2008). Although selection rates in the VIS/NIR studies appear similar to those from broader wavelength hyperspectral studies there are some notable differences. The initial peak in selection rates present in both sets is shifted towards shorter blue wavelengths, and a greater importance of the red edge over the red minimum is evident for the VIS/NIR studies. However, the overall pattern is the same with two peaks in the rate of selection at both the blue/green and red reflectance minima, with yellow wavelength bands having the lowest selection rate save for the sub 400 nm bands that appear in a very limited number of studies. Although the VIS/NIR studies do not cover the full NIR region, selection rates for the red edge and shorter wavelength NIR are closely matched

between both groups (Figure 1). The overall higher rates present in the VIS/NIR table results from the smaller number of studies in that group, with selection rates tending to decrease as more studies are added. Additionally, the relatively small number of studies included in the VIS/NIR group prevents the analysis of specific subsets such as canopy and leaf. The following discussion of selection rates refers to VIS/SWIR studies (Table 2) and is generally applicable to the VIS/NIR studies, although, particular discussion of the VIS/NIR studies is included when required.

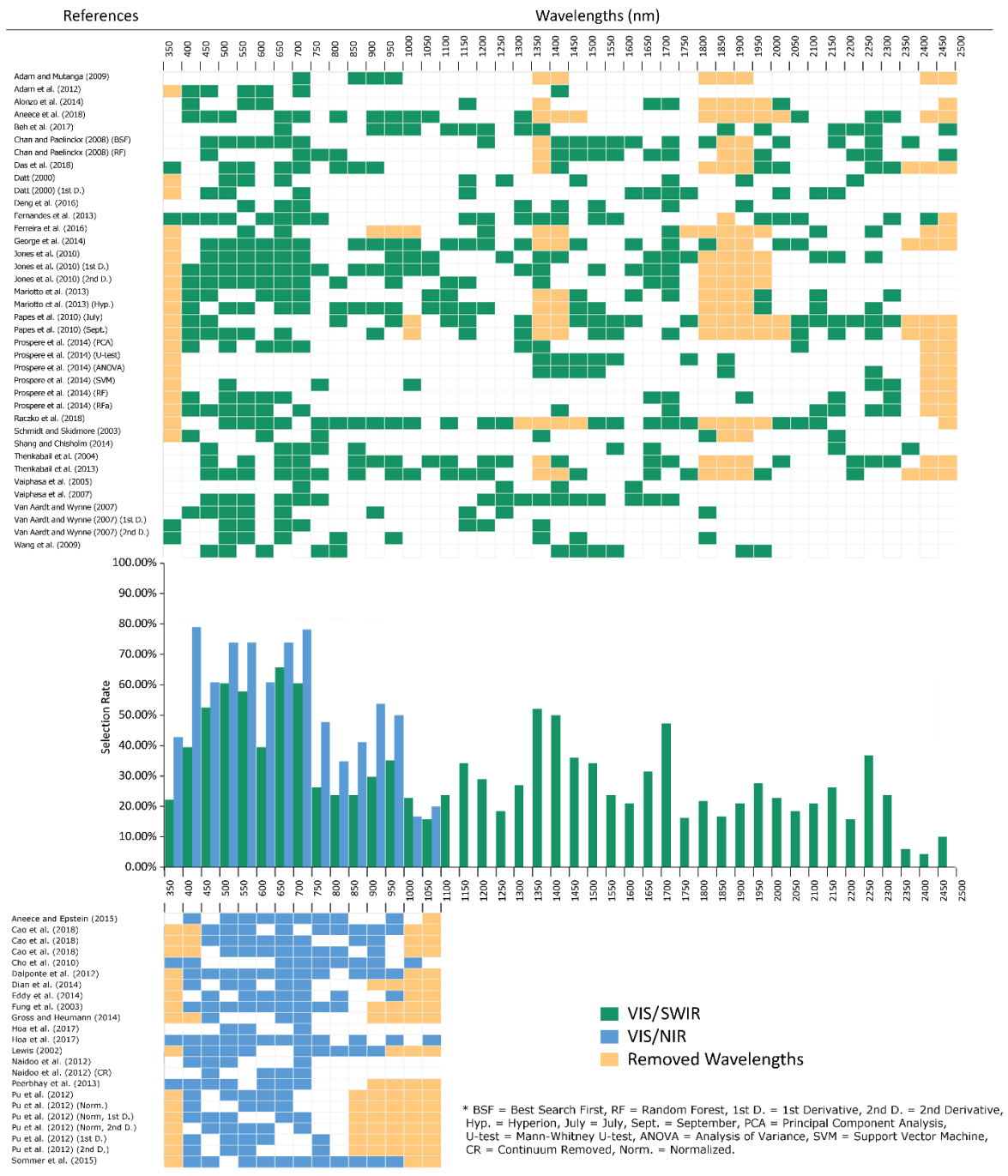


Figure 1. Waveband selection binned at 50 nm intervals for the VIS/SWIR studies (350 -2500 nm) green, VIS/NIR studies (350 – 1100 nm) blue. Orange filled cells represent waveband regions removed from a study due to noise. Selection rate is the percentage of studies that selected a given 50 nm region for species classification. Each row of the table is an individual study, with each column being a 50 nm range bin. Green/blue shaded bins represent at least one waveband being selected from within that range, orange shaded bins represent removed wavelength regions (e.g. major water absorption regions). Wavelength bins were only removed if the entire 50 nm region was removed due to noise/atmospheric effects in that particular study.

2.2.2. Visible (VIS; 400 – 700 nm)

Primarily a region of low reflectance in living foliage, typically as low as 5 – 10% with the exception of the green peak at ~550 nm where reflectance can be more than twice that of surrounding wavelengths (Figure 2). Reflection in the visible wavelengths is dominated by absorptions from foliar pigments. Differences in leaf pigments between species have been identified by many studies as important factors for discrimination (Fernandes, Aguiar et al. 2013), despite variability in the VIS being generally low compared to longer wavelengths (Asner 1998, Ollinger 2011). Of the pigments, chlorophyll a and b have the strongest influence over absorption in this region, followed by that of carotenoids, and anthocyanins whose effects are predominantly masked by that of chlorophyll. The visible region is one of the most influential regions for classification, with the vast majority of studies in this review selecting bands from within it. The visible wavelengths can be divided into three regions of high discriminatory value spanning almost the entire visible range; the blue/blue-green edge (400 – 499 nm), the green peak centred around 550 nm, and the red reflectance minimum (650 – 700 nm) (Figure 2). Of these, the red reflectance minimum, specifically bands near 680 nm has previously been identified as the most commonly selected and critical band centre for crop type discrimination (Thenkabail, Smith et al. 1999). The continued selection of 680 nm, along with neighbouring bands in later studies has validated the importance of this region amongst agricultural crop studies (Thenkabail, Smith et al. 2002, Thenkabail, Enclona et al. 2004, Mariotto, Thenkabail et al. 2013, Eddy, Smith et al. 2014), as well as for other vegetation types (Fung, Yan Ma et al. 2003, Galvão, Roberts et al. 2009, Cho, Debba et al. 2010, Naidoo, Cho et al. 2012, Ferreira, Grondona et al. 2013, Peerbhay, Mutanga et al. 2013, Fassnacht, Neumann et al. 2014, Gross and Heumann 2014, Shang and Chisholm 2014, Richter, Reu et al. 2016). In addition to the obvious relationship with chlorophyll, absorption in the red region has been related to anthocyanin content, a foliar pigment responsible for the red colouration in leaves (Blackburn 2006), particularly evident in juvenile leaves of certain species (Peerbhay, Mutanga et al. 2013).

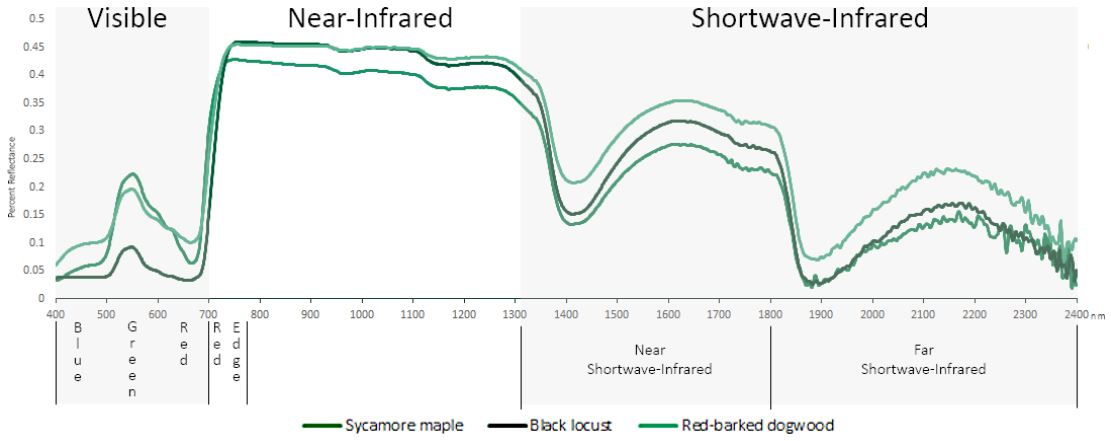


Figure 2. Example hyperspectral reflectance of 3 species of tree and key broad regions of the electromagnetic spectrum (400 – 2400 nm).

The green region has the second highest selection rate amongst both the VIS and entire measured spectrum (Figure 1). Wavebands selected in this region tend to be focused around the green reflectance peak at approximately 550 nm, which is strongly correlated with chlorophyll content (Thomas and Gausman 1977). The green peak, either manually chosen as a spectral variable as a representation of chlorophyll content or selected via feature selection, has demonstrated importance in classifying species (Castro-Esau, Sánchez-Azofeifa et al. 2006, Pu 2009, Jones, Coops et al. 2010, Peerbhay, Mutanga et al. 2013, Alonzo, Bookhagen et al. 2014). Additionally, absorption in wavebands within the green region adjacent to the reflectance peak is associated with xanthophylls and anthocyanins. Xanthophyll pigments protect against photo-oxidation of the photosynthesis reaction centres during high light conditions (Demmig-Adams and Adams 1996), resulting in short term changes in reflectance at 531 nm. This band, along with 570 nm, makes up the photochemical reflectance index (Gamon, Penuelas et al. 1992). Anthocyanins can be estimated by an index using anthocyanin's absorption maximum near 550 nm, and a band from the red edge, usually 700 nm (Gitelson, Merzlyak et al. 2001). Although not necessarily associated with these additional pigments, studies have selected bands along the leading edge of the green reflectance peak between 500 – 550 nm (Cho, Debba et al. 2010).

Selection from the blue region (400 – 449 nm) has the third highest rate in the VIS region, though the blue-green edge (450-499 nm) has an almost equal rate of selection to the green region (55.8% and 58.8% respectively). The importance of blue bands has

been established for discriminating within groups of conifers and between conifers and broadleaf species (Gong, Pu et al. 1997, van Aardt 2000), though its inclusion in approximately half of the studies, many of which include non-coniferous species, indicates its importance in general for a wider range of vegetation types. Some of these non-coniferous studies focused on the savanna ecosystem where blue bands along with the red reflectance minimum and red edge were informative (Cho, Debba et al. 2010, Naidoo, Cho et al. 2012). Blue wavelengths are strongly influenced by chlorophyll absorption, along with carotenoid absorption features present in the 450 – 499 nm region. Carotenoids have proven important for discrimination of senescent leaves, when decay of chlorophyll and the diminishing of the strong chlorophyll-absorption feature reveal the carotenoid absorption feature (Richter, Reu et al. 2016).

However, studies have noted that strong similarities between the visible reflectance of different species can decrease the significance of VIS wavelengths for classification purposes. In one such study the NIR region was more informative at distinguishing species than the VIS, with spectral differences in the VIS region being non-significant between species (Karlovska, Grīnfelde et al. 2016). Additionally, in a study of tropical trees, (Rivard, Sanchez-Azofeifa et al. 2008) performed feature selection and classification on various datasets derived from the same original spectra. One dataset included the wavelengths 350 – 2500 nm, another excluded the VIS, while another excluded the SWIR. Although it was found that the full spectrum produced greater overall classification accuracy, and both reduced datasets produced lower overall accuracies, individual accuracies for certain species remained high. The classification model excluding the VIS region maintained high accuracies for 6 out of 20 species, whereas the model excluding the SWIR maintained high accuracies for 5 out 20 species. Although the importance of the VIS region has been described by many authors and is clearly seen in the binned data, studies such as (Rivard, Sanchez-Azofeifa et al. 2008) demonstrate that wavelength importance is dependent on the species included in the study.

2.2.3. Red Edge (680-780 nm)

The red edge encompasses the region from the red reflectance minimum around 680 nm to the NIR shoulder at approximately 780 nm, and indicates the sharp increase in

reflectance from the VIS to NIR regions associated with strong chlorophyll absorptions and internal leaf structure (Figure 2). The inflection point of the slope in this region has been defined as the Red Edge Position (REP) (Clevers, De Jong et al. 2002), and its strong correlation with chlorophyll concentration has seen it used as an indicator of stress and senescence in vegetation (Dawson and Curran 1998, Gholizadeh, Mišurec et al. 2016). In the VIS-SWIR studies the red edge region as represented by the 700-749 nm bin has the same rate of selection as the red minimum bin, whereas the VIS-NIR studies have a slightly higher red edge rate than red minimum. However, as previously stated, the delineation between the red minimum bin (650-699nm) and the red edge bin (700 – 749nm) means that bands selected from the lower point of the red edge would be included in the red minimum bin, potentially skewing red edge band selection rates.

The red edge region has been described as one of the most informative and frequently selected regions in a number of studies, where the authors have attributed its importance to its correlation with chlorophyll abundance, nitrogen concentration, water content and structural features such as Leaf Area Index (LAI) (Clark, Roberts et al. 2005, Dalponte, Bruzzone et al. 2009, Cho, Debba et al. 2010, Dalponte, Bruzzone et al. 2012). Additionally, significant variation of the red edge region between species has been documented after a first derivative transformation has been applied to the spectra (Cochrane 2000). The red edge has proven especially important in studies discriminating species with high levels of chlorophyll and high LAI values such as the giant reed (*Arundo donax*), in which a distinctive “red shift” is seen where the REP is located at higher wavelengths (Adam and Mutanga 2009, Fernandes, Aguiar et al. 2013). This “red shift” mirrors the “blue shift” of the REP where its position is shifted towards the shorter blue wavelengths associated with a decrease in chlorophyll and used to monitor senescence or stress (Rock, Hoshizaki et al. 1988).

2.2.4. Near Infrared (NIR) (700-1327 nm)

The NIR is often defined to include wavelengths within the red edge region (680 -780 nm) (Jones, Coops et al. 2010): as this region has been previously discussed, this section focusses on the NIR plateau (780 – 1327 nm). The high reflectance of the plateau results from scattering of photons within the leaf structure due to a change in the refractive index from liquid water to air within the inter-cellular spaces (Knipling 1970). Two minor

water absorption features at ~ 980 nm and ~ 1200 nm are the only major features of plateau. Along with water content, the depth and width of these absorptions can be influenced by the spectral recording method. Canopy scale spectra tend to produce deeper and wider absorption features compared to leaf scale, at which absorption features can vary with leaf stack thickness (Clark, Roberts et al. 2005). High levels of intraspecific variability have been identified in the NIR and related to leaf age, water and chlorophyll concentration, as well as herbivory, necrosis and epiphyll cover (Datt 2000, Clark, Roberts et al. 2005). Wavebands selected in studies reporting these high levels of intraspecific variation have generally been limited to the water absorption features (Datt 2000, Dalponte, Bruzzone et al. 2012), although it has been suggested to avoid band selection from within or near water absorption features due to this high level of within class variability, specifically for Eucalypts (Kumar 2007, Kumar, Skidmore et al. 2010, Shang and Chisholm 2014). Despite this, (Clark, Roberts et al. 2005) reported greater interspecific variability in the NIR, particularly at the canopy scale, potentially related to species specific photon scattering caused by differences in canopy architecture, a result also reported by other studies (van Aardt 2000, Karlovska, Grīnfelde et al. 2016). However, it has been suggested that the importance of the NIR and SWIR in (Clark, Roberts et al. 2005) is linked to the time delay between leaf collection and spectral measurement, causing a decrease in water content and affecting waveband importance (Ferreira, Grondona et al. 2013).

Even when the high selection rate of the red edge is included, the average selection rate of the NIR is close to being half of that of the VIS, placing it third after the near SWIR. However, there are two small peaks in the rate of selection within the NIR, in bins 950 – 999 nm and 1150 – 1199 nm, both of which are associated with water absorption features near 980 and 1200 nm. Despite having one of the lowest rates, some studies have reported that bands in the NIR plateau are the most strongly discriminating (Schmidt and Skidmore 2003, Wang, Xu et al. 2009).

2.2.5. Shortwave Infrared (SWIR) (1328 – 2500 nm)

Based on the binned results (Figure 1) the SWIR can be divided into two distinct regions, the near SWIR (NSWIR) from 1350 – 1800 nm including the strong water absorption feature at 1350 – 1450 nm, and the far SWIR (FSWIR) from 1800 -2500 nm, including

another strong water absorption feature from 1800 – 2000 nm. The wavebands associated with these water absorption features that mark the start of the SWIR and separate the near and far SWIR are often removed from spectra due to high levels of noise, as are the bands at the far end of the SWIR above 2400 nm. Selection rates within the NSWIR is on average the second highest, primarily caused by high rates of selection at 1350 – 1450 and 1700 – 1750 nm. This initial high selection rate spanning two consecutive bins is associated with the water absorption feature focused around 1400 nm. However, these bins are often removed in studies, primarily when hyperspectral imagery is used due to increased noise that is not as prevalent in lab or field spectra. Selection rates then drop in the mid-NSWIR bands before peaking again for the 1700-1750 nm bin, containing wavebands often associated with lignin, cellulose, tannins, and other biochemical constituents of foliar and non-foliar plant matter (Curran 1989, Elvidge 1990). The FSWIR has the lowest average band selection rate with its highest selection at bin 2250 – 2299 nm most likely associated with weak absorption features of cellulose and lignin present at 2270nm (Curran 1989, Elvidge 1990).

As the selection results suggest, wavebands selected from the SWIR are reported in the literature as being associated with water absorption (Datt 2000, Thenkabail, Enclona et al. 2004, Adam, Mutanga et al. 2012, Ferreira, Grondona et al. 2013, Thenkabail, Mariotto et al. 2013, Fassnacht, Neumann et al. 2014, Shang and Chisholm 2014, Ferreira, Zortea et al. 2016) or the weak harmonic and overtone absorptions from biochemicals such as lignin, starch, and cellulose (Thenkabail, Enclona et al. 2004, Wang, Xu et al. 2009, Jones, Coops et al. 2010, Ferreira, Grondona et al. 2013, Thenkabail, Mariotto et al. 2013, Alonzo, Bookhagen et al. 2014, Shang and Chisholm 2014, Lehmann, Große-Stoltenberg et al. 2015, Ferreira, Zortea et al. 2016). However, as described in regards to the NIR, selection of bands in or near water absorption features may not be suitable for classification in field or lab spectra, due to high levels of intraspecific variance (Kumar, Skidmore et al. 2010, Shang and Chisholm 2014). Additionally, bands selected from leaf scale spectra in the two major water absorption features would not be applicable to remotely sensed imagery as they coincide with low irradiance levels resulting from atmospheric water absorption. The observation of higher selection rates in the NSWIR compared to the FSWIR has previously been made with

studies noting the importance of NSWIR bands and absence of selection from the FSWIR (Jones, Coops et al. 2010, Alonzo, Bookhagen et al. 2014), even when visual differences between species were apparent (Wang, Xu et al. 2009). Possible reasons for this reduced selection of the FSWIR could be high levels of LAI or leaf water content masking the biochemical features present in this region (Kokaly, Asner et al. 2009), or high correlation between the FSWIR, NSWIR and VIS bands (Alonzo, Bookhagen et al. 2014).

2.2.6. Canopy and Leaf Scale Spectral Selection Rates

The red edge has been demonstrated as one of the most frequently selected regions (Figure 1), though the remainder of the NIR (consisting of 12 bins from 750 – 1349 nm) has the second lowest mean selection rate, only slightly higher than the FSWIR. As the literature has identified an increase in importance of the NIR for canopy spectra, a comparison of band selection rates for each bin was made between canopy and leaf scale spectral studies (Figure 3). Leaf spectra were defined as only containing pure leaf reflectance, with canopy being primarily leaf spectra, though also containing non-photosynthetic vegetation and potentially background reflectance. This comparison shows a clear increase in selection rates for the NIR bins associated with water absorption features for the canopy studies, and a related decrease amongst the leaf scale spectra. Differences are also apparent in the visible regions with a substantial increase in selection of the leading edge of the green peak, and a decrease in selection of the trailing edge of the green peak for leaf scale studies compared to canopy level (Figure 3). This would indicate a blue-shift for green bands selected in leaf scale spectra, and a red-shift of selected bands for canopy spectra. Differences in spectral reflectance for the VIS region have been identified at different scales, with branch/canopy spectra including reflectance characteristics from non-foliar sources, shadows and uneven lighting, as well as generally displaying an increase in pigment absorption features (Asner 1998, Clark, Roberts et al. 2005). Variation in selection rates is also evident in the SWIR, most notably a broad region of increased selection for canopy spectra across four bins from 1950 to 2149 nm, and a sudden peak at 1800 – 1850 nm. The selection peaks of the canopy spectra correspond to regions of water absorption which have demonstrated an increase in depth and width in canopy studies. However, the disparity between canopy and leaf scale spectra is potentially exaggerated by the fact that a majority of canopy

studies eliminate these wavebands due to noise concerns, with the remaining few studies selecting these wavebands as being discriminatory. Increased selection of the broader region could also be related to water absorption, as well as structural components such as lignin and cellulose, particularly from non-photosynthetic material in the canopy (Clark, Roberts et al. 2005). The NSWIR however demonstrates the highest degree of conformity for a large region, covering 9 bins from 1300 – 1750 nm.

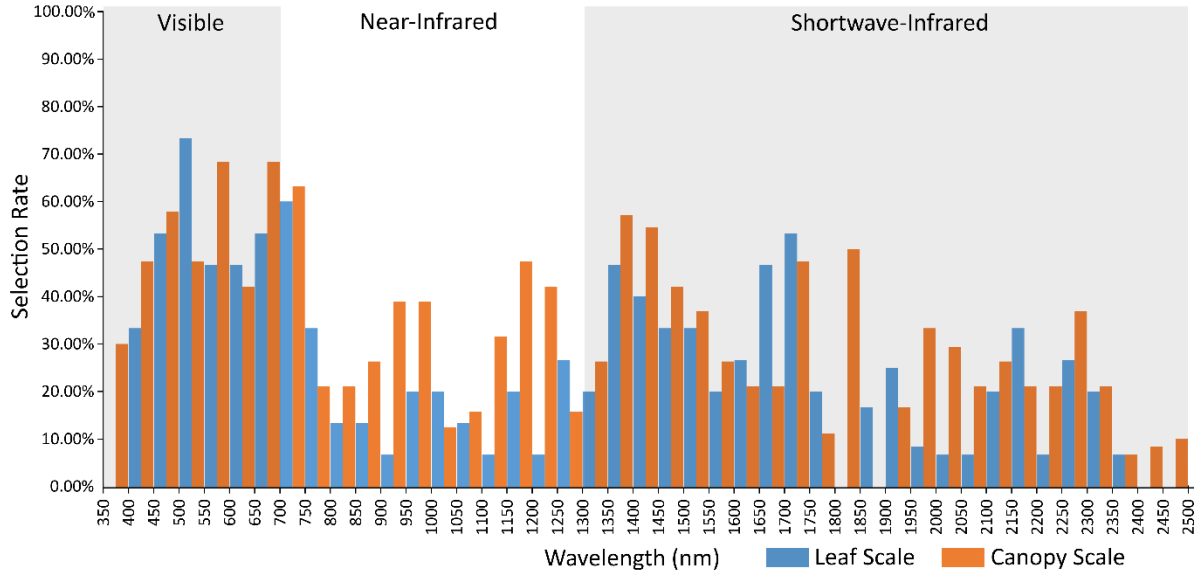


Figure 3. Waveband selection rates for 350 – 2500 nm studies (Table 2) per 50 nm bins of both canopy and leaf scale spectra.

2.3. Feature selection

Feature selection is implemented to select a subset of features to improve generalization and computation requirements while preserving or improving classification accuracy. In this review feature selection and waveband selection is used interchangeably. Feature selection techniques are generally divided into three categories: filter, wrapper, and embedded methods. Filter methods are named as such as they act as a pre-processing step that filters out irrelevant features. Filter methods are known to be computationally fast and efficient, though are generally outperformed by the other methods, as well as not being able to handle nonlinear relationships (Alonso-Atienza, Rojo-Álvarez et al. 2012).

2.3.1. Filter methods

Analysis of variance (ANOVA) is a parametric statistical filter method to determine

significant differences between group means. Related to ANOVA is the non-parametric Mann-Whitney U-test, and the Kruskal-Wallis test which extends the Mann-Whitney U-test for more than two groups (Schmidt and Skidmore 2003). Following initial dimensionality reduction by one of these methods a secondary feature selection step to further reduce the number of selected features is used, such as LDA (Wang, Xu et al. 2009), Classification and Regression Trees (CART) (Adam and Mutanga 2009, Fernandes, Aguiar et al. 2013), or manual selection of known influential bands (Schmidt and Skidmore 2003, Prospere, McLaren et al. 2014, Shang and Chisholm 2014). This secondary selection step found important bands in the VIS and SWIR, with reduced selection of NIR bands (Wang, Xu et al. 2009, Fernandes, Aguiar et al. 2013), however, the reverse was found by (Adam and Mutanga 2009) where CART secondary selection was restricted to NIR wavelengths. The remainder of the studies manually selected bands that differentiated the greatest number of species pairs (Prospere, McLaren et al. 2014), or selected known influential bands from the wavelengths that demonstrated high levels of pairwise group variance (Schmidt and Skidmore 2003, Shang and Chisholm 2014).

2.3.2. Wrapper methods

Wrapper methods search for a subset of features that gives the best classification performance, with the best performing subset being selected. Although generally considered to outperform filter methods, wrappers are known to be computationally demanding and can suffer from overfitting (Alonso-Atienza, Rojo-Álvarez et al. 2012).

Two of the studies reviewed implemented Genetic Algorithms (GA), in which wavebands are encoded as genes that are subsequently grouped into chromosomes. These chromosomes are allowed to evolve over many generations where their fitness, as determined by a classifier, controls their likelihood to reproduce and pass their genes onto the next generation. Fitness of chromosomes is determined each generation by a chosen classifier, with the classification accuracy of each chromosome being its fitness score, chromosomes with increased fitness are more likely to reproduce. Both studies used the same dataset of lab measured tropical mangrove leaves (Vaiphasa, Ongsomwang et al. 2005, Vaiphasa, Skidmore et al. 2007). Selection of bands differed

between the two studies despite the use of the same dataset and feature selector, though methodologies did differ. The variability of selected bands with similar classification performance seen between these studies demonstrates that multiple band selections can perform classification equally well. The ensemble of chromosomes used in (Vaiphasa, Skidmore et al. 2007) helped identify key regions for discriminating target species related to biophysical and biochemical aspects of the vegetation that may have been missed if a study was reliant upon the first single chromosome to reach the stopping criterion. This is apparent when comparing the bands selected in both studies, with (Vaiphasa, Ongsomwang et al. 2005) selecting no VIS bands, resulting in the authors concluding that pigments were not significant for the discrimination of the target species. However, the importance of the VIS, particularly the green region became apparent in (Vaiphasa, Skidmore et al. 2007) where 21 out of 120 total bands were selected from 513 +/- 19 nm.

Forward feature selection (FFS) is a wrapper method of feature selection that begins with a model containing a single feature that best discriminates the classes, with new features iteratively added to the model based on their ability to improve class discrimination (Pudil, Novovičová et al. 1994). FFS was implemented by (Gross and Heumann 2014) in their comparison between floral and leaf spectra: only the results for leaf spectra are discussed here. The leaf spectra within this study were constrained to 475 – 900 nm at 1 nm increments, with only 8 wavebands being selected. These bands came from narrow regions of the spectra, occurring in the blue, red minimum and red edge regions. In a similar spectral range of 402.9 to 989.1 nm of airborne collected spectra a very different feature selection trend was observed by (Dalponte, Bruzzone et al. 2012) following use of the FFS variant Sequential Floating Feature Selection (SFFS). Wavebands were selected from across the entire reduced spectrum, with a notable gap in selection occurring in the NIR between 800-849 nm. Selection differences exhibited between these studies could be related to the differences in target species, leaf or canopy scale spectra, or version of FFS used. The only VIS-SWIR study in this review to use FFS applied it to AVIRIS imagery of urban street trees (Alonzo, Bookhagen et al. 2014). However, feature selection was only performed to identify spectral regions responsible for species separability, with all bands used for classification. These

informative spectral regions matched a number of known informative regions from the literature, such as water absorption in the NIR, cellulose and lignin features in the SWIR, and bands associated with photosynthetic pigments in the VIS. Interestingly however, the highly selected red minimum and red edge were not selected in this study, along with the majority of the NIR.

2.3.3. Embedded methods

Despite being described as a wrapper method in (Prospere, McLaren et al. 2014), Recursive Feature Elimination with a Support Vector Machine (SVM-RFE) is considered to be an embedded method (Guyon, Weston et al. 2002). Embedded methods differ from wrappers, as they do not treat the classifier as a black box, rather, features are selected using information gained whilst training the classifier (Deng and Runger 2012). A claimed strength of SVMs as a classifier is its reported independence of the Hughes effect or curse of dimensionality (Melgani and Bruzzone 2004, Pal and Mather 2004). However, it has been shown that SVM classifications can be affected by the Hughes effect and can benefit from dimensionality reduction of its inputs, especially when sample sizes are small (Pal and Foody 2010).

In order to be used as a feature selection method, (Prospere, McLaren et al. 2014) implemented Recursive Feature Elimination (RFE) with a SVM, determining that from the original 401 bands the optimal number of features to include for classification is 20, after 1-5, 10, 15, 20 and 30 where all evaluated. The 20 bands selected demonstrated a number of trends that were not apparent in the other feature selection methods implemented in the same study. Firstly, the bands formed four distinct contiguous clusters at 520 – 530 nm, 745 – 775 nm, 1005 – 1030 nm, 2295 – 2305 nm, and then a final single band at 2345 nm at 5 nm bandwidths. Secondly, the wavelengths of certain selected bands were also unique amongst the methods used, with SVM-RFE being the only method to select bands from the NIR plateau out of all feature selection methods implemented in (Prospere, McLaren et al. 2014). Additionally, being the only method to not select bands from the NSWIR. Although not reported in a manner suitable for inclusion in Table 1, (Fassnacht, Neumann et al. 2014) also performed feature ranking with a SVM. As with (Prospere, McLaren et al. 2014), (Fassnacht, Neumann et al. 2014) identified the optimal number of features to be between 15 and 20, depending on the

dataset, pre-processing, and feature selection methods used. Unlike (Prospere, McLaren et al. 2014) where the SVM selected bands from distinct contiguous regions, (Fassnacht, Neumann et al. 2014) report the SVM selecting bands evenly spread over the entire spectrum.

Random Forest (RF) is an ensemble classification method, in which a number of decision tree classifiers are trained from a sub-sample of the dataset, with their results combined via a voting system. One third of samples are retained for validation purposes known as the out-of-bag (OOB) samples with the remaining in-the-bag samples being used to construct the decision tree (Breiman 2001).

Of the original 72 bands in (Naidoo, Cho et al. 2012) between 384.8 nm and 1054.3 nm, 8 were selected for classification via RF. Although no other feature selection method was implemented in this study, a previous study by (Cho, Debba et al. 2010) performed feature selection with the Spectral Angle Mapper (SAM) add-on Selector using the same data. This resulted in the selection of a far greater 31 bands. Upon binning of the bands at 50 nm a clear difference in the selection methods are evident (Figure 1). The RF selected bands of (Naidoo, Cho et al. 2012) are focused in the 400 – 550 nm region with a single band from the red edge at 706 nm, whereas the SAM bands are focused along the red edge and NIR plateau between 650 and 950 nm, with additional bands in the 350 – 450 and 1000 – 1050 nm regions.

As with the bands selected in Naidoo, Cho et al. (2012), the RF selected bands in Chan and Paelinckx (2008) fell within 4 bins in the VIS and VNIR regions. However, in Naidoo, Cho et al. (2012) band selection was focused on the green region with limited selection apparent in the red and NIR plateau with the exception of a single band near the Red Edge inflection point. This focus was seemingly switched in Chan and Paelinckx (2008) with bands falling into the bins along the Red Edge up to the NIR plateau shoulder, with the remaining bin occurring at the blue/green edge. The Chan and Paelinckx (2008) study also offers a comparison to an alternative feature selection method using the best-first search (BFS) algorithm as a wrapper. The band selection techniques differ greatly in the VIS and VNIR regions with only the bins at 450-499 and 700-749 in common.

However, band selection is more similar at longer wavelengths where the majority of bands were selected by both methods.

The wavebands selected via RF in Prospere, McLaren et al. (2014) are in direct opposition to those selected by RF in Chan and Paelinckx (2008). Selected bands in Chan and Paelinckx (2008) mainly occurred along the red edge and NIR plateau shoulder, no band was selected in this region by Prospere, McLaren et al. (2014). Instead, focus was placed on the green, yellow, and red regions of the VIS wavelengths, an area completely ignored by Chan and Paelinckx (2008) RF selector, though significant for their BFS selection. Additionally, Prospere, McLaren et al. (2014) provided the top 20 informative bands determined by a RF classifier using the full 201 waveband dataset. Although these two implementations of RF differed in selecting bands, the overall trend was very similar with high selection rates in the VIS, low in the NIR and similar selection throughout the SWIR.

Additionally a study by Adam, Mutanga et al. (2012) produced waveband selections similar to those in Prospere, McLaren et al. (2014) with similar results in the VIS with the exception of no selection in the early green (500 – 549 nm), and selection of the red edge bin rather than the red minimum. The biggest difference between Adam, Mutanga et al. (2012) and all other RF studies is the reduced selection at longer wavelengths, although all studies essentially ignored the NIR, only selected 2 bands from the SWIR, both within the same NSWIR bin at the water absorption feature near 1400 – 1449 nm.

2.3.4. Comparison of Stepwise Discriminant Analysis (SDA) with non-SDA feature selectors

Stepwise discriminant analysis is a filter method that selects a subset of features by attempting to minimise within-class variation while simultaneously maximising between-class variation (Huberty 1994). Although a number of metrics are available to determine class separability, Wilk's lambda is by far the most frequently used to enter and remove variables from the selection in a stepwise manner. Some studies reported Wilk's lambda approaching zero and becoming asymptotic, indicating near perfect separation of classes (Thenkabail, Mariotto et al. 2013). Features selected after this point can be safely removed from the model as they will not substantially increase classification accuracy.

This normally resulted in the selection of 10 – 20 wavebands (Datt 2000, Thenkabail, Enclona et al. 2004, Van Aardt and Wynne 2007, Mariotto, Thenkabail et al. 2013, Thenkabail, Mariotto et al. 2013).

SDA in general selects wavebands more uniformly across the spectrum than other methods, though the greatest number of selected bands is still found in the VIS (Figure 4). The most significant difference for selection rates is the increased importance of the NIR beyond the red edge. The NIR demonstrates significant selection with the use of SDA in all bar a first derivative dataset from Van Aardt and Wynne (2007), and Datt (2000), with the author of the latter suggesting high levels of intraspecific variance due to differences in leaf maturity as the reason no bands were selected in this region.

Upon comparing the selection rates of SDA studies compared to non-SDA, a clear difference in selection of NIR bands is apparent. As with the difference between canopy and leaf scale spectra the increased selection is focused around the NIR water absorption features (Figure 4). Additionally, in the VIS there is significantly higher selection for the blue, green and red regions in SDA studies. In order to determine if spectral acquisition scale or feature selection technique had greater influence on band selection the selection rates were further subset into canopy studies using SDA and non-SDA feature selection and leaf scale studies using SDA and non-SDA selection (Figure 5). It is apparent that the feature selection method has a greater impact on band selection rates, with SDA selecting from the NIR with far greater rates than the non-SDA methods in both canopy and leaf scale studies. The non-SDA methods demonstrated minimal selection in the NIR beyond the red edge for leaf scale spectra, with only a slight increase in selection for canopy spectra focused around the water absorption wavelengths from 1150-1250 nm. The studies that did select from the NIR with leaf scale samples via non-SDA methods stated that the selected bands represented differences in internal reflectance for leaf scale spectra (Vaiphasa, Skidmore et al. 2007). The blue and red shifts around the green peak for canopy and leaf scale spectra are still evident once the data has been subset into SDA/non-SDA, although it becomes apparent that the high rates of selection in many parts of the VIS is driven by the SDA studies. However, the use of SDA does not explain the selection rates of the VIS for the reduced spectral domain VIS/NIR studies, as only a single study used SDA for feature selection, perhaps indicating

an alternate driving force. The red edge demonstrates its robustness to variations in measurement scale and band selection technique as it was frequently selected for all study subsets, although slightly less frequently for leaf scale spectra with non-SDA feature selection.

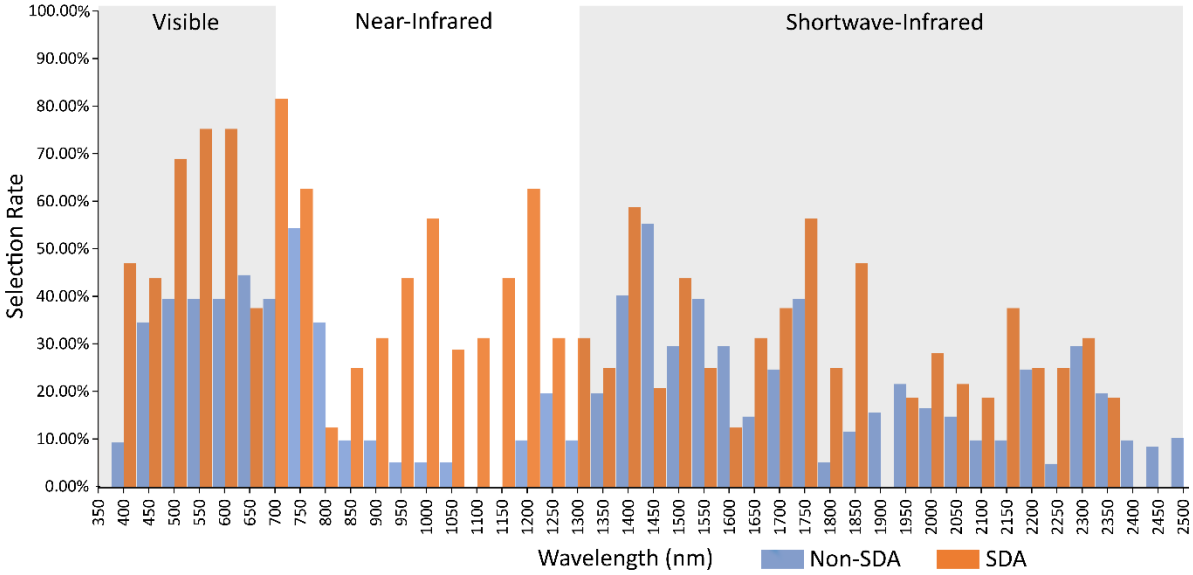


Figure 4. Waveband selection rates for 350 – 2500 nm studies that used SDA feature selection, and selection rate of all other feature selection methods combined.

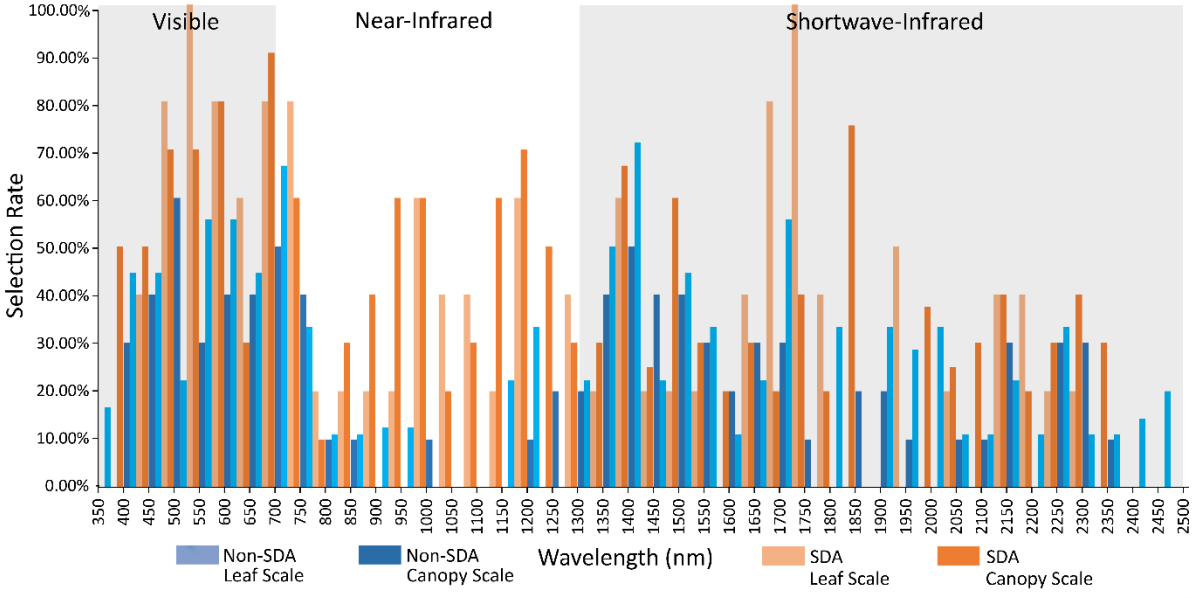


Figure 5. Waveband selection rates for 350 – 2500 nm studies that used SDA feature selection subset by canopy and leaf scale spectra, and selection rate of all other feature selection methods combined subset by canopy and leaf scale spectra.

According to Thompson (1989) “Stepwise analytic methods may be among the most popular research practices employed in both substantive and validity research”. Despite this statement being made in the late 1980s the use of SDA in approximately a third of the studies included in this review demonstrates its continued popularity, being by far the most used method encountered. However, the widespread use of stepwise methods

has prompted strong arguments against its usage (Huberty 1994, Thompson 1995, Whitaker 1997, Flom and Cassell 2007), particularly so when utilised in a predictive discriminant analysis application such as feature selection for classification (Huberty and Barton 1989). The studies that utilised SDA in this review made no mention of these criticisms and therefore no direct attempt to mitigate them. Despite this, Eddy, Smith et al. (2014) did validate their model with 20 repetitions of 1000 random samples, with the final feature subset being based on the selection rates of features across the repetitions, consideration of important features identified in the literature from Lewis (2002) and Thenkabail, Enclona et al. (2004) as well as the results from Principal Component Analysis (PCA). PCA is a mathematical transformation used to produce uncorrelated features from the spectral features, reducing dimensionality whilst retaining the most informative spectral data. Additionally, Thenkabail, Enclona et al. (2004), and Mariotto, Thenkabail et al. (2013), included SDA as part of an ensemble of feature selection methods, again determining the final feature subset based on the selection rates of features across all methods within the ensemble. Although one of these ensemble methods (Lambda – Lambda plots) allows for the identification and removal of correlated features, in both cases it was run in parallel to SDA with removal of correlated features occurring after features had been selected. The remaining studies reported no efforts to mitigate the concerns of using SDA for feature selection (Datt 2000, Van Aardt and Wynne 2007, Jones, Coops et al. 2010, Papeş, Tupayachi et al. 2010, Thenkabail, Mariotto et al. 2013, George, Padalia et al. 2014).

It must be acknowledged that the sub-setting of reviewed studies into canopy and leaf scale, and then into SDA and non-SDA meant each class was only represented by a small number of samples (~8 per class), though leaf-SDA was only represented by 5 studies extracted from 2 papers. As a result of this, a few outliers are evident, such as the 100% selection in bin 1700 - 1749 nm and the 100% selection of the 500 - 549 nm bin, both associated with the low leaf-SDA sample size. Additionally, the comparison of SDA to non-SDA may disguise selection biases of the non-SDA methods as they are often only represented by one or two studies, with any bias they may exhibit being masked by the selection rates of the other methods.

2.4. Study design influence

All aspects of a study design influence waveband selection. However, many of these aspects may be outside the control or be heavily constrained for the researcher, such as target classes, number of samples and collection method, though the researcher often has control over data pre-processing, feature selection, and classification methods. Due to this, and the apparent influence of feature selectors previously described we focus on how choice of feature selection method effects waveband selection.

In order to ascertain any influence feature selection may have over waveband selection, some of the most common feature selection methods were applied to a synthesised dataset. A key requirement for these experiments is the need for a dataset with many species with large numbers of samples, something generally lacking in vegetation hyperspectral data. To accomplish this a hyperspectral synthesis method was created (Hennessy, Clarke et al. 2021) to allow for the creation of any number of samples from 22 species of New Zealand plants. The synthesised dataset consisted of 500 samples per class with 540 wavebands from 350 – 2450 nm at 3 nm bandwidths, excluding regions of high noise.

Three experiments were devised. First, each feature selection method was performed on the same dataset cross-validated 10 times (eg. rf_0, rf_1 … rf_9) selecting the top 30 discriminative wavebands, revealing any possible biases in waveband selection resulting from the choice of feature selection method (Figure 6, 7). Secondly, feature selection was performed on datasets consisting of different classes and samples to simulate many different studies, giving an idea if attributes of the samples affects the wavebands being selected, which will impact generalizability and transferability. Variants of this experiment were performed were the classes used remained the same as did the number of samples, though actual samples were randomly selected, additionally the same classes with differing numbers of samples. Results for these variants did not significantly differ and therefore aren't shown here.

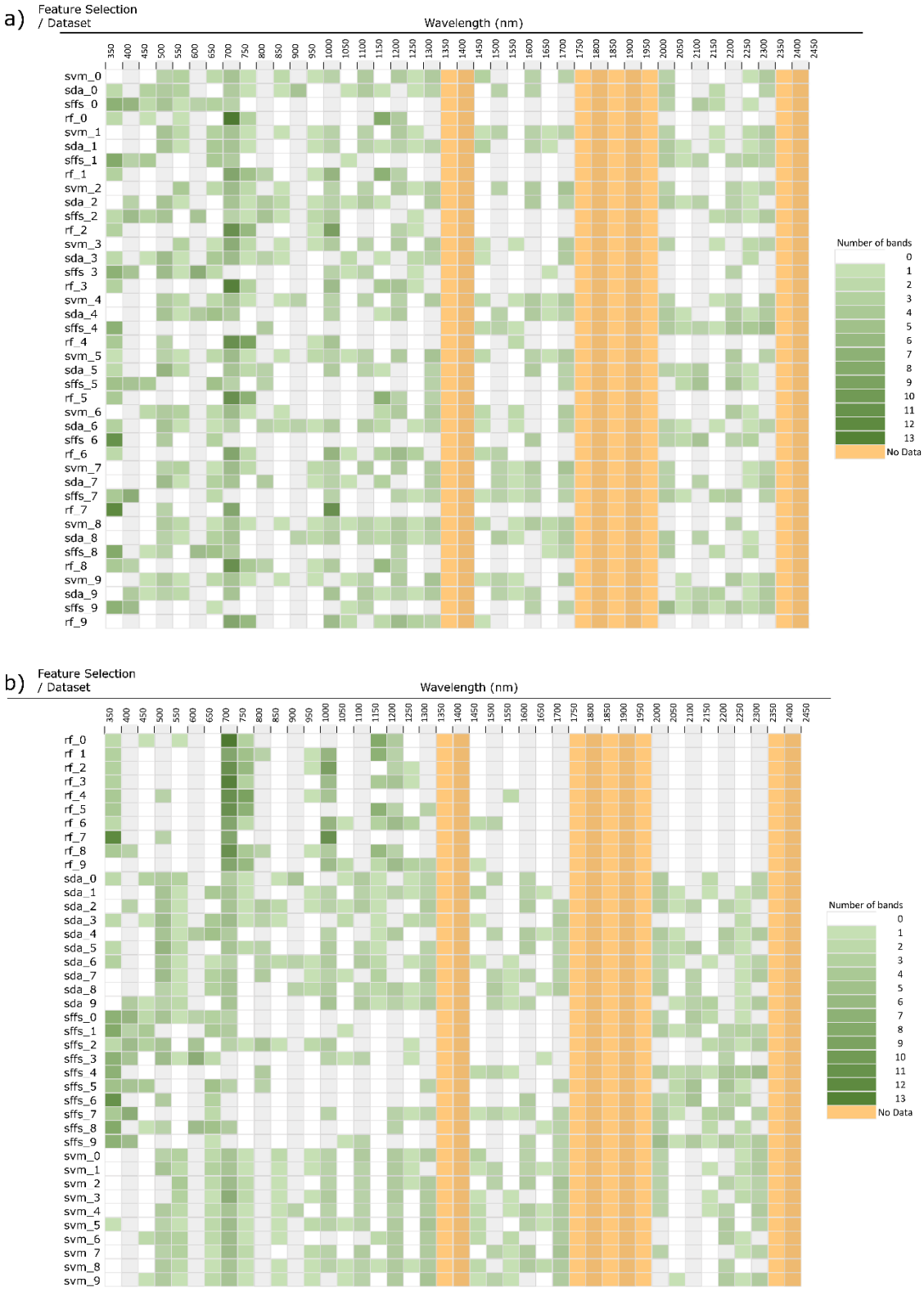


Figure 6. a) Histogram of band feature selection binned at 50nm, ordered by dataset. Four feature selectors run on the same dataset 10x cross-validation (new dataset consisting of 10 classes and 200 samples for each cross-val.). b) Results of Fig. 6 an ordered by feature selection method. (RF = Random Forest, SDA = Stepwise Discriminant Analysis, SFFS = Sequential Floating Feature Selection, SVM = Support Vector Machine)

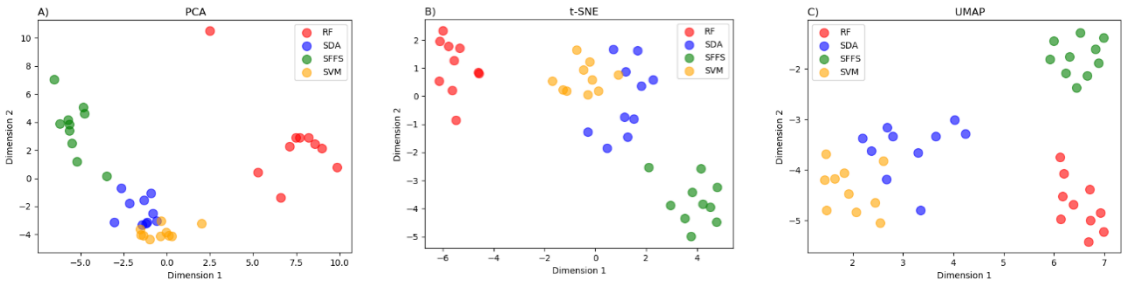


Figure 7. a) PCA dimensional reduction of histogram waveband feature selection. b) T-SNE dimensional reduction of histogram waveband feature selection. c) UMAP dimensional reduction of histogram waveband feature selection.

Each dataset produced significantly different waveband selections. This is especially evident in (Figure 6b) where the histogram is ordered by feature selector, placing each repetition with a new dataset next to each other. Here it is clear the RF favours the red edge and NIR bands, essentially ignoring the SWIR. SFFS has higher selection in the VIS, especially at shorter wavelengths, minimal selection in the NIR and medium selection in late SWIR. SDA and SVM are the most similar due to both selecting broadly and relatively evenly along the entire spectrum. Dimensionality reduction techniques offer a way to visualize the relationship between selection histograms (Figure 7a,b,c). Due to their broad general selection, SDA and SVM are grouped close to each other with SFFS and RF adjacent though separate. Further, the histograms are clearly grouped by feature selection method rather than by dataset, indicating that feature selection method is a dominant factor as to which wavebands are selected.

2.5. Conclusion

This review of hyperspectral vegetation classification literature has determined that every aspect of a study can greatly influence selected wavebands and classification performance. However, despite this we have identified some important and consistent patterns that appear throughout the literature. Visible wavelengths and their associations with photosynthetic pigments have played important discriminatory roles in a wide range of studies, their high levels of selection clearly evident in this review (Figure 1). Selection rates in the VIS showed only minor variations between VIS/SWIR studies and the VIS/NIR (Figure 1), although the comparisons between canopy and leaf scale spectra demonstrated significant differences (Figure 3). The discriminatory value of the red edge has been well documented with its close relationship to chlorophyll

concentration and structural features. This is reflected in the consistently high rates of selection of the red edge as well as the robustness of its selection with only minor variation in magnitude between the comparisons. The inclusion of structural features in canopy spectra can provide high levels of interspecific variation in the NIR, primarily in the form of differences in albedo, rather than spectral shape (van Aardt 2000). However, selection rates from the non-red edge NIR are low, with the selected bands generally being related to water absorption features and potentially high levels of within class variability. Additionally, the NIR has demonstrated the greatest degree of variability between the canopy and leaf scale spectra studies. Wavebands selected in the SWIR are associated with water absorption and non-photosynthetic biochemicals, with selection rates heavily skewed towards the NSWIR over the FSWIR.

The reported importance of NIR bands (Schmidt and Skidmore 2003, Wang, Xu et al. 2009) seems to be contentious, primarily being driven by the use of a single feature selection technique. Comparisons between selection rates for the NIR with and without the use of SDA as the feature selector are starkly contrasted, with NIR importance being significantly higher with the use of SDA. The criticisms of SDA and stepwise methods in general perhaps offer an answer to the selection biases presented in this review.

It is clearly apparent that there is no single best feature selection method, with the same method performing very differently within and between studies. This suggests that either multiple methods should be applied to the data or an ensemble of multiple methods may be the best practice, a conclusion recognized by this review, previously suggested by some studies (Chan and Paelinckx 2008). Additionally, multiple subsets of selected features have proven to discriminate species equally well (Prospere, McLaren et al. 2014), or alternatively, no feature selection, with the original data outperforming feature selected subsets (Chan and Paelinckx 2008, Alonzo, Bookhagen et al. 2014, Sommer, Holzwarth et al. 2015). Additionally, as computation power, dataset sizes, and machine learning techniques all increase, the need for feature selection as a data reduction technique becomes less necessary.

This review has established that the variability in waveband selection seen between studies, driven by study parameters beyond characteristics of the target samples

prevents determination of generalizable high utility spectral regions for specific taxonomic discrimination. Broad trends such as the importance of VIS and red edge wavelengths are apparent, independent of plant groupings, though in and of themselves they are not sufficiently specific for taxonomic discrimination. The possibility of discriminatory spectral regions being associated with specific taxonomic, structural or functional groupings of plants is inconclusive due to the large degree of variability apparent between studies. This is further highlighted by the apparent dominance of feature selector choice over other parameters for waveband selection (Figure 6, 7).

2.6. References

Adam, E. and O. Mutanga (2009). "Spectral discrimination of papyrus vegetation (Cyperus papyrus L.) in swamp wetlands using field spectrometry." *ISPRS Journal of Photogrammetry and Remote Sensing* 64(6): 612-620.

Adam, E., O. Mutanga, D. Rugege and R. Ismail (2012). "Discriminating the papyrus vegetation (Cyperus papyrus L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP." *International Journal of Remote Sensing* 33(2): 552-569.

Alonso-Atienza, F., J. L. Rojo-Álvarez, A. Rosado-Muñoz, J. J. Vinagre, A. García-Alberola and G. Camps-Valls (2012). "Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection." *Expert Systems with Applications* 39(2): 1956-1967.

Alonzo, M., B. Bookhagen and D. A. Roberts (2014). "Urban tree species mapping using hyperspectral and lidar data fusion." *Remote Sensing of Environment* 148: 70-83.

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

Blackburn, G. A. (2006). "Hyperspectral remote sensing of plant pigments." *Journal of Experimental Botany* 58(4): 855-867.

Breiman, L. (2001). "Random forests." *Machine Learning* 45(1): 5-32.

Castro-Esau, K. L., G. A. Sánchez-Azofeifa, B. Rivard, S. J. Wright and M. Quesada (2006). "Variability in leaf optical properties of Mesoamerican trees and the potential for species classification." *American Journal of Botany* 93(4): 517-530.

Chan, J. C.-W. and D. Paelinckx (2008). "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery." *Remote Sensing of Environment* 112(6): 2999-3011.

Cho, M. A., P. Debba, R. Mathieu, L. Naidoo, J. Van Aardt and G. P. Asner (2010). "Improving discrimination of savanna tree species through a multiple-endmember spectral angle mapper approach: Canopy-level analysis." *IEEE Transactions on Geoscience and Remote Sensing* 48(11): 4133-4142.

Clark, M. L., D. A. Roberts and D. B. Clark (2005). "Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales." *Remote Sensing of Environment* 96(3): 375-398.

Clevers, J., S. De Jong, G. Epema, F. Van Der Meer, W. Bakker, A. Skidmore and K. Scholte (2002). "Derivation of the red edge index using the MERIS standard band setting." *International Journal of Remote Sensing* 23(16): 3169-3184.

Cochrane, M. (2000). "Using vegetation reflectance variability for species level classification of hyperspectral data." *International Journal of Remote Sensing* 21(10): 2075-2087.

Curran, P. J. (1989). "Remote sensing of foliar chemistry." *Remote Sensing of Environment* 30(3): 271-278.

Dalponte, M., L. Bruzzone and D. Gianelle (2012). "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data." *Remote Sensing of Environment* 123: 258-270.

Dalponte, M., L. Bruzzone, L. Vescovo and D. Gianelle (2009). "The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas." *Remote Sensing of Environment* 113(11): 2345-2355.

Datt, B. (2000). “Recognition of eucalyptus forest species using hyperspectral reflectance data.” Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International, IEEE.

Dawson, T. and P. Curran (1998). "Technical note A new technique for interpolating the reflectance red edge position." *International Journal of Remote Sensing* 19(11): 2133–2139.

Demmig-Adams, B. and W. W. Adams (1996). "The role of xanthophyll cycle carotenoids in the protection of photosynthesis." *Trends in Plant science* 1(1): 21-26.

Deng, H. and G. Runger (2012). Feature selection via regularized trees. Neural Networks (IJCNN), The 2012 International Joint Conference on, IEEE.

Eddy, P., A. Smith, B. Hill, D. Peddle, C. Coburn and R. Blackshaw (2014). "Weed and crop discrimination using hyperspectral image data and reduced bandsets." *Canadian Journal of Remote Sensing* 39(06): 481-490.

Elvidge, C. D. (1990). "Visible and near infrared reflectance characteristics of dry plant materials." *Remote Sensing* 11(10): 1775-1795.

Fassnacht, F. E., H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, C. Straub and A. Ghosh (2016). "Review of studies on tree species classification from remotely sensed data." *Remote Sensing of Environment* 186: 64-87.

Fassnacht, F. E., C. Neumann, M. Förster, H. Buddenbaum, A. Ghosh, A. Clasen, P. K. Joshi and B. Koch (2014). "Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6): 2547-2561.

Fernandes, M. R., F. C. Aguiar, J. M. Silva, M. T. Ferreira and J. M. Pereira (2013). "Spectral discrimination of giant reed (Arundo donax L.): A seasonal study in riparian areas." *ISPRS Journal of Photogrammetry and Remote Sensing* 80: 80-90.

Ferreira, M. P., A. E. B. Grondona, S. B. A. Rolim and Y. E. Shimabukuro (2013). "Analyzing the spectral variability of tropical tree species using hyperspectral feature selection and leaf optical modeling." *Journal of Applied Remote Sensing* 7(1): 073502-073502.

Ferreira, M. P., M. Zortea, D. C. Zanotta, Y. E. Shimabukuro and C. R. de Souza Filho (2016). "Mapping tree species in tropical seasonal semi-deciduous forests with hyperspectral and multispectral data." *Remote Sensing of Environment* 179: 66-78.

Flom, P. L. and D. L. Cassell (2007). Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use. NorthEast SAS Users Group Inc 20th Annual Conference: 11-14th November 2007; Baltimore, Maryland.

Fung, T., H. F. Yan Ma and W. L. Siu (2003). "Band selection using hyperspectral data of subtropical tree species." *Geocarto International* 18(4): 3-11.

Galvão, L. S., D. A. Roberts, A. R. Formaggio, I. Numata and F. M. Breunig (2009). "View angle effects on the discrimination of soybean varieties and on the relationships between vegetation indices and yield using off-nadir Hyperion data." *Remote Sensing of Environment* 113(4): 846-856.

Gamon, J., J. Penuelas and C. Field (1992). "A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency." *Remote Sensing of environment* 41(1): 35-44.

George, R., H. Padalia and S. Kushwaha (2014). "Forest tree species discrimination in western Himalaya using EO-1 Hyperion." *International Journal of Applied Earth Observation and Geoinformation* 28: 140-149.

Gholizadeh, A., J. Mišurec, V. Kopačková, C. Mielke and C. Rogass (2016). "Assessment of Red-Edge Position Extraction Techniques: A Case Study for Norway Spruce Forests Using HyMap and Simulated Sentinel-2 Data." *Forests* 7(10): 226.

Gitelson, A. A., M. N. Merzlyak and O. B. Chivkunova (2001). "Optical properties and nondestructive estimation of anthocyanin content in plant leaves." *Photochemistry and Photobiology* 74(1): 38-45.

Gong, P., R. Pu and B. Yu (1997). "Conifer species recognition: An exploratory analysis of in situ hyperspectral data." *Remote Sensing of Environment* 62(2): 189-200.

Gross, J. W. and B. W. Heumann (2014). "Can flowers provide better spectral discrimination between herbaceous wetland species than leaves?" *Remote Sensing Letters* 5(10): 892-901.

Guyon, I., J. Weston, S. Barnhill and V. Vapnik (2002). "Gene selection for cancer classification using support vector machines." *Machine Learning* 46(1): 389-422.

Hennessy, A., K. Clarke and M. Lewis (2021). "Generative adversarial network synthesis of hyperspectral vegetation data." *Remote Sensing* 13(12): 2243.

Huberty, C. J. (1994). Applied discriminant analysis. New York Wiley

Huberty, C. J. and R. M. Barton (1989). "An Introduction to Discriminant Analysis." *Measurement and Evaluation in Counseling and Development* 22(3): 158-168.

Jones, T. G., N. C. Coops and T. Sharma (2010). "Employing ground-based spectroscopy for tree-species differentiation in the Gulf Islands National Park Reserve." *International Journal of Remote Sensing* 31(4): 1121-1127.

Karlovska, A., I. Grīnfelde, I. Alsiņa, G. Priedītis and D. Roze (2016). “Plant reflected spectra depending on biological characteristics and growth conditions.” International Scientific Conference Rural Development 2017.

Knipling, E. B. (1970). "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation." *Remote Sensing of Environment* 1(3): 155-159.

Kokaly, R. F., G. P. Asner, S. V. Ollinger, M. E. Martin and C. A. Wessman (2009). "Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies." *Remote Sensing of Environment* 113: S78-S91.

Kumar, L. (2007). "A comparison of reflectance characteristics of some Australian eucalyptus species based on high spectral resolution data—discriminating using the visible and NIR regions." *Journal of Spatial Science* 52(2): 51-64.

Kumar, L., A. K. Skidmore and O. Mutanga (2010). "Leaf level experiments to discriminate between eucalyptus species using high spectral resolution reflectance data: use of derivatives, ratios and vegetation indices." *Geocarto International* 25(4): 327-344.

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

Lewis, M. (2002). "Spectral characterization of Australian arid zone plants." *Canadian Journal of Remote Sensing* 28(2): 219-230.

Mariotto, I., P. S. Thenkabail, A. Huete, E. T. Slonecker and A. Platonov (2013). "Hyperspectral versus multispectral crop-productivity modeling and type discrimination for the HyspIRI mission." *Remote Sensing of Environment* 139: 291-305.

Melgani, F. and L. Bruzzone (2004). "Classification of hyperspectral remote sensing images with support vector machines." *IEEE Transactions on Geoscience and Remote Sensing* 42(8): 1778-1790.

Naidoo, L., M. A. Cho, R. Mathieu and G. Asner (2012). "Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment." *ISPRS Journal of Photogrammetry and Remote Sensing* 69: 167-179.

Ollinger, S. V. (2011). "Sources of variability in canopy reflectance and the convergent properties of plants." *New Phytology* 189(2): 375-394.

Pal, M. and G. M. Foody (2010). "Feature selection for classification of hyperspectral data by SVM." *IEEE Transactions on Geoscience and Remote Sensing* 48(5): 2297-2307.

Pal, M. and P. M. Mather (2004). "Assessment of the effectiveness of support vector machines for hyperspectral data." *Future Generation Computer Systems* 20(7): 1215-1225.

Papeş, M., R. Tupayachi, P. Martinez, A. Peterson and G. Powell (2010). "Using hyperspectral satellite imagery for regional inventories: a test with tropical emergent trees in the Amazon Basin." *Journal of Vegetation Science* 21(2): 342-354.

Peerbhay, K. Y., O. Mutanga and R. Ismail (2013). "Commercial tree species discrimination using airborne AISA Eagle hyperspectral imagery and partial least squares discriminant analysis (PLS-DA) in KwaZulu–Natal, South Africa." *ISPRS Journal of Photogrammetry and Remote Sensing* 79: 19-28.

Prospere, K., K. McLaren and B. Wilson (2014). "Plant species discrimination in a tropical wetland using in situ hyperspectral data." *Remote Sensing* 6(9): 8494-8523.

Pu, R. (2009). "Broadleaf species recognition with in situ hyperspectral data." *International Journal of Remote Sensing* 30(11): 2759-2779.

Pudil, P., J. Novovičová and J. Kittler (1994). "Floating search methods in feature selection." *Pattern Recognition Letters* 15(11): 1119-1125.

Richter, R., B. Reu, C. Wirth, D. Doktor and M. Vohland (2016). "The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area." *International Journal of Applied Earth Observation and Geoinformation* 52: 464-474.

Rivard, B., G. Sanchez-Azofeifa, S. Foley and J. Calvo-Alvarado (2008). "Species classification of tropical tree leaf reflectance and dependence on selection of spectral bands." *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*: 141-159.

Rock, B., T. Hoshizaki and J. Miller (1988). "Comparison of in situ and airborne spectral measurements of the blue shift associated with forest decline." *Remote Sensing of Environment* 24(1): 109-127.

Schmidt, K. and A. Skidmore (2003). "Spectral discrimination of vegetation types in a coastal wetland." *Remote Sensing of Environment* 85(1): 92-108.

Shang, X. and L. A. Chisholm (2014). "Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(6): 2481-2489.

Sommer, C., S. Holzwarth, U. Heiden, M. Heurich, J. Müller and W. Mauser (2015). "Feature based tree species classification using hyperspectral and lidar data in the Bavarian Forest National Park." EARSel eProceedings, Vol. 14, Special Issue 2: 9th EARSeL Imaging Spectroscopy Workshop, 2015.

Thenkabail, P., R. Smith and E. De Pauw (1999). "Hyperspectral vegetation indices for determining agricultural crop characteristics, CEO research publication series No. 1." *Center for Earth Observation, Yale University Press, New Haven*.

Thenkabail, P. S., E. A. Enclona, M. S. Ashton and B. Van Der Meer (2004). "Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications." *Remote Sensing of Environment* 91(3): 354-376.

Thenkabail, P. S., I. Mariotto, M. K. Gumma, E. M. Middleton, D. R. Landis and K. F. Huemmrich (2013). "Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/EO-1 data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6(2): 427-439.

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

Thomas, J. and H. Gausman (1977). "Leaf reflectance vs. leaf chlorophyll and carotenoid concentrations for eight crops." *Agronomy Journal* 69(5): 799-802.

Thompson, B. (1989). "WHY WONT STEPWISE METHODS DIE", Amer Counseling Association 5999 Stevenson Ave, Alexandria, VA 22304-3300.

Thompson, B. (1995). "Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial", Sage Publications Sage CA: Thousand Oaks, CA.

Vaiphasa, C., S. Ongsomwang, T. Vaiphasa and A. K. Skidmore (2005). "Tropical mangrove species discrimination using hyperspectral data: A laboratory study." *Estuarine, Coastal and Shelf Science* 65(1): 371-379.

Vaiphasa, C., A. K. Skidmore, W. F. de Boer and T. Vaiphasa (2007). "A hyperspectral band selector for plant species discrimination." *ISPRS Journal of Photogrammetry and Remote Sensing* 62(3): 225-235.

Van Aardt, J. and R. Wynne (2007). "Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results." *International Journal of Remote Sensing* 28(2): 431-436.

van Aardt, J. A., R.H. Wynne (2000). "Spectral separability among six southern tree species." Photogrammetric Engineering and Remote Sensing, 67 (12) (2001), pp. 1367-137

Wang, J., R. Xu and S. Yang (2009). "Estimation of plant water content by spectral absorption features centered at 1,450 nm and 1,940 nm regions." *Environmental Monitoring and Assessment* 157(1-4): 459-469.

Wang, J., R. Xu and S. Yang (2009). "Estimation of plant water content by spectral absorption features centered at 1,450 nm and 1,940 nm regions." *Environmental Monitoring and Assessment* 157(1): 459-469.

Whitaker, J. S. (1997). "Use of Stepwise Methodology in Discriminant Analysis." Paper presented at the annual meeting of the Southwest Educational Research Association, Austin.

CHAPTER THREE: Generative adversarial network synthesis of hyperspectral vegetation data

This chapter is published as:

Hennessy, A., K. Clarke and M. Lewis (2021). "Generative adversarial network synthesis of hyperspectral vegetation data." *Remote Sensing* 13(12): 2243.

This chapter appears as published with minor modifications to reference style and formatting of figure and table captions for consistency with the remainder of the thesis

Statement of Authorship

Title of Paper	Generative adversarial network synthesis of hyperspectral vegetation data	
Publication Status	☒ Published	☐ Accepted for Publication
	☐ Submitted for Publication	☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Hennessy, A., K. Clarke and M. Lewis (2021). "Generative adversarial network synthesis of hyperspectral vegetation data." *Remote Sensing* 13(12): 2243.	

Principal Author

Name of Principal Author (Candidate)	Andrew Hennessy	
Contribution to the paper	Conceptualization, Methodology, data curation, formal analysis, writing – original draft preparation, writing – review and editing.	
Overall percentage (%)	70%	
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature		Date: 13/07/2021

Co-author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kenneth D. Clarke	
Contribution to the paper	Methodology, formal analysis, writing – review and editing	
Signature		Date: 13/07/2021

Name of Co-Author	Megan M. Lewis	
Contribution to the paper	Methodology, formal analysis, writing – review and editing	
Signature		Date: 12/07/2021

Abstract

New, accurate and generalisable methods are required to transform the ever-increasing amount of raw hyperspectral data into actionable knowledge for applications such as environmental monitoring and precision agriculture. Here we apply advances in generative deep learning models to produce realistic synthetic hyperspectral vegetation data, whilst maintaining class relationships. Specifically, a Generative adversarial network (GAN) is trained using the Cramér distance on two vegetation hyperspectral datasets, demonstrating the ability to approximate the distribution of the training samples.

Evaluation of the synthetic spectra shows that they respect many of the statistical properties of the real spectra, conforming well to the sampled distributions of all real classes. Creation of an augmented dataset consisting of synthetic and original samples was used to train multiple classifiers with increases in classification accuracy seen under almost all circumstances. Both datasets showed improvements in classification accuracy ranging from a modest 0.16 % for the Indian Pines set and a substantial increase of 7.0 % for the New Zealand vegetation. Selection of synthetic samples from sparse or outlying regions of the feature space of real spectral classes demonstrated increased discriminatory power over those from more central portions of the distributions.

3.1. Introduction

Hyperspectral (HS) Earth observation has increased in popularity in recent years, driven by advancements in sensing technologies, increased data availability, research and institutional knowledge. The big data revolution of the 2000s and significant advances in data processing and machine learning (ML) have seen hyperspectral approaches used in a broad spectrum of applications, with methods of data acquisition covering wide-ranging spatial and temporal resolutions.

For researchers aiming to classify or evaluate vegetation, hyperspectral remote sensing offers rich spectral information detailing the influences of pigments, biochemistry, structure and water absorption whilst having the benefits of being non-destructive, rapid, and repeatable. These phenological variations imprint a sort of 'spectral fingerprint' that allows hyperspectral data to differentiate vegetation at taxonomic units ranging from broad ecological types to species and cultivars (Hennessy, Clarke et al. 2020). Acquiring labelled hyperspectral measurements of vegetation is expensive and time-consuming, resulting in limited training datasets for supervised classification techniques. However, this has been slightly alleviated through multi/hyperspectral data-sharing portals such as ECOSTRESS (JPL/NASA 2021) and SPECCHIO (Hueni, Chisholm et al. 2020). Supervised classification of such high dimensional data has had to rely on feature reduction or selection techniques in order to overcome small training sample sizes and avoid the curse of dimensionality, also called the 'Hughes phenomenon'. Additionally, the general requirement of large training datasets in ML has meant limited success has been had when trying to leverage recent ML progress towards classification of HS data, often leading to overfitting of models and poor generalizability.

Data augmentation (DA), the process of artificially increasing training sample size, has been implemented by the ML community when the problem of small or imbalanced datasets has been encountered. DA methods vary from simple pre-processing steps such as mirroring, rotating or scaling of images (Taylor and Nitschke 2017), to more complicated simulations (Goodenough and Brown 2017, Wang 2019) and generative models (Bissoto, Perez et al. 2018, Wang, Kang et al. 2018). DA for time series or 1D data consists of the addition of noise, or methods such as time dilation, or cut & paste

(Wen, Sun et al. 2020). However, when dealing with non-spatial HS data these methods would be unsuitable, as it is important to maintain reflectance and waveband relationships in order to ensure class labels are preserved. Methods of DA such as physics-based models (Jacquemoud, Verhoef et al. 2009), or noise injection (Slavkovikj, Verstockt et al. 2015, Nalepa, Myller et al. 2019) have been applied to HS data. Whilst successful, these methods are either simplifications of reality and require domain-dependent knowledge of target features in the case of physical models or rely upon random noise, potentially producing samples that only approximate the true distribution.

Generative Adversarial Networks (GANs) have been used successfully in many fields as a DA technique, often for images, timeseries/1D (Harada, Hayashi et al. 2019), sound synthesis (Donahue, McAuley et al. 2018), or anonymising medical data (Esteban, Hyland et al. 2017). GANs consist of two neural networks trained in an adversarial manner. The generator (G) network produces synthetic copies mimicking the real training data while the discriminator (D) network attempts to identify if a sample was from the real dataset or produced by G. The D is scored on its accuracy in identifying real from synthetic data, before passing feedback to G allowing it to learn how best to fool D and improve generation of synthetic samples (Goodfellow, Pouget-Abadie et al. 2014).

The use of GANs to generate synthetic HS data is a relatively new field of study. GANs of varying architectures ranging from 1D spectral (Audebert, Le Saux et al. 2018, Zhan , Hu et al. 2017, Xu, Du et al. 2018) to 2D (Feng, Yu et al. 2019), and 3D spectral-spatial (Zhu, Chen et al. 2018) with differing data embeddings including individual spectra, HS images, and principal components have been examined. All have been able to demonstrate the ability to generate synthesized hyperspectral data and to improve classification outcomes to varying degrees whether through DA or conversion of the GANs discriminator model to a classifier. However, issues such as training instability and mode collapse, a common form of overfitting are prevalent.

The work presented in this paper applies advances in generative models to overcome limitations previously encountered by (Audebert, Le Saux et al. 2018) to produce more

realistic synthetic HS vegetation data and eliminate reliance on PCA to reduce dimensionality and stabilise training. Specifically, we train a GAN using the Cramér distance on two vegetation HS datasets, demonstrating the ability to approximate the distribution of the training samples while encountering no evidence of mode collapse. We go on to demonstrate the use of these synthetic samples for data augmentation and reducing under sampling of class distributions.

3.2. Generative adversarial networks - Background

GANs are a type of generative machine learning algorithm known as an implicit density model. This type of model does not directly estimate or fit the data distribution but rather generates its own data which is used to update the model. Since first being introduced by (Goodfellow, Pouget-Abadie et al. 2014) GANs have become a dominant field of study within ML/DL, with numerous variants and being described as “the most interesting idea in the last 10 years in machine learning” by leading AI researcher (LeCun 2016). Although sometimes utilizing non-neural network architectures, GANs generally consist of two neural networks, sometimes more, that compete against each other in a minimax game. This is where one neural network, the discriminator, attempts to reduce its “cost” or error as much as possible. This occurs in an adversarial manner, where the discriminator is trained to maximize the probability of correctly labelling whether a sample originates from the original data distribution or has been produced by the generator. Simultaneously the generator is trained to minimize the probability that the discriminator correctly labels the sample (Gui, Sun et al. 2020).

Because of this the training of GANs is notoriously unstable, with issues such as the discriminator’s cost quickly becoming zero and providing no gradient to update the generator, or the generator converging onto a small subset of samples that regularly fool the discriminator, a common issue known as mode collapse. Considerable research has gone into attempting to alleviate these issues, improve training stability and improve quality of synthetic samples, so much so that during 2018 more than one GAN-related paper was being released every hour (Gui, Sun et al. 2020).

Unlike non-adversarial neural networks, the loss function of a GAN doesn't converge to an optimal state, making the loss values meaningless in respect to evaluating the

performance of the model. In an attempt to alleviate this problem the Wasserstein GAN (WGAN) was developed to use the Wasserstein distance, also known as the Earth Mover's (EM) distance which results in an informative loss function for both D and G that converges to a minimum (Arjovsky, Chintala et al. 2017). Rather than the D having sigmoid activation in its final layer producing a binary classification of real or fake, WGAN approximates the Wasserstein distance which is a regression task detailing the distance between the real and fake distributions. Due to gradient loss being a common weakness with WGANs they were improved by applying weight clipping to the losses with a gradient penalty (GP) (Gulrajani, Ahmed et al. 2017) further improving training stability.

3.3. Experimental design

Here we implement the CramérGAN, a GAN variant using the Cramér/energy distance as the Ds loss, reportedly offering improved training stability and increased generative diversity over WGANs (Bellemare, Danihelka et al. 2017). This choice was informed by our preliminary testing of wGAN and wGAN-GP that produced noisy synthesized samples and lower standard deviations, in addition to the learning instability and poor convergence previously reported for wGAN, which may explain mode collapse encountered by (Audebert, Le Saux et al. 2018).

Individual models were trained for each hyperspectral class, for a total of 38 models. Each model was trained for 50,000 epochs, at a ratio of 5:1 (5 training iterations of D for every 1 of G) using the Adam optimizer at a learning rate of 0.0001, with beta1 = 0.5 and beta2 = 0.9. The latent noise vector was generated from a normal distribution with length 100. The G consists of two fully connected dense layers followed by two convolution layers, all using the ReLu activation function save for the final convolution layer using Sigmoid activation. The final layer of G reshaped the output to be a 2D array with shape (batch size * number of bands). A similar architecture was used for the D, though reversed. Starting with two convolution layers into a flatten layer, followed by 2 fully connected dense layers, all layers of D used Leaky ReLu activation except the final layer which used a linear function (Figure 8).

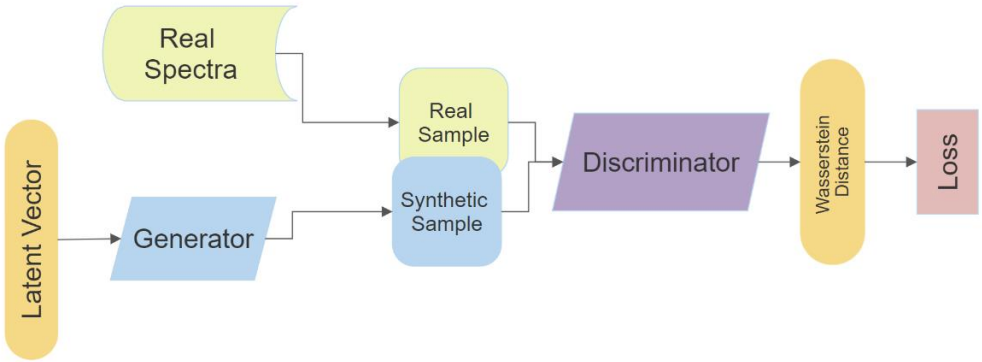


Figure 8. Schematic of GAN architecture

The three classification models (SVM, RF and NN) were evaluated in four permutations: trained on real data and evaluated on real data (real – real); trained on real data and evaluated on synthetic data (real – synthetic); trained on synthetic data and evaluated on synthetic data (synthetic– synthetic); and trained on synthetic data and evaluated on real data (synthetic– real). Each dataset was split into training and testing subsets with 10 times cross-validation. All synthetic datasets were restricted to the same number of samples per class as the real datasets unless specified otherwise. The real - real experiments were expected to have the highest accuracy and offer a baseline of comparison for the synthetic samples. If the accuracy of real – synthetic is significantly higher than real - real it potentially indicates the generator has not fully learned the true distribution of the training samples. Conversely, accuracy being significantly lower could mean the synthetic samples are outside the true distribution and are an unrealistic representation of the spectra.

Extending this analysis, synthetic-synthetic and synthetic-real experiments were performed with the number of synthesized training samples increasing from 10 to 490 samples by increments of 10 samples per class. The real – synthetic and real – real were included for comparison with a consistent number of training samples, though training and evaluation subsets differed every iteration. The initial DA experiment was performed with the same number of samples for real and synthetic datasets with the augmented set having twice the number before the number of synthetic samples were incremented by 10 from 10 to 490 samples per class.

The data augmentation capabilities of the synthetic spectra were evaluated by similar methods. First the three classifiers were trained with either real, synthetic or both combined into an augmented dataset and tested against an evaluation dataset that was not used in the training of the GAN.

All code was written and executed in Python 3.7. The CramérGAN based upon (Song 2017) using the Tensorflow 1.8 framework. Support Vector Machine (SVM), and Random Forests (RF) classifiers make use of the Scikit-Learn 0.22.2 library, with Tensorflow 1.8 utilised for the neural network (NN) classifier. Additionally, Scikit-Learn 0.22.2 provided the dimensionality reduction functions for Principal Components Analysis (PCA) and t-distributed Stochastic Neighbourhood Embedding (t-SNE), with Uniform Manifold Approximation and Projection (UMAP) being a standalone library. Hyperparameters for all functions are provided in Appendix A.

Potential classification power of a sample was estimated with the C metric devised in (Mountrakis and Xi 2013) for the purpose of predicting the likelihood of correctly classifying an unknown sample by measuring its Euclidean distance in feature space to samples in the training dataset. (Mountrakis and Xi 2013) demonstrated a strong correlation between close proximity to number of training samples and likelihood of correctly being classified. The C metric is bound between -1 indicating low likelihood and 1 high likelihood of successful classification.

Rather than focusing on the proximity of an unknown sample to a classifier's training data we are interested in the distance of each synthesized sample to that of the real data in order to evaluate any potential increase in information density. We hypothesise that a C value closer to the lower bound for a synthetic sample would indicate it being further away from real data points and of any synthetic samples with C values close to the upper bound. Such a sample could potentially contain greater discriminatory power for the classifier as it essentially fills a gap in feature space of the class distribution.

To determine whether some samples of the NZ dataset provide more information to the classifier than others, and that the improvement in classification accuracy is not purely from increased sample size, the distance of each generated sample was measured to all real samples of its class before being converted to a C value as per

(Mountrakis and Xi 2013), with an *h* value range of 1-50 at increments of 1. Two data subsets were then created using the first 100 spectral samples after all synthetic samples were ordered by their C value in ascending (most distant) and descending (least distant) order. The first 100 samples from each ordered dataset rather than the full 500 were used to maximize differences, reduce computation time and simplify figures.

3.4. Datasets

Two hyperspectral datasets were used to train the GAN: Indian Pines agricultural land cover types (INDI); and New Zealand plant spectra (NZ). The Indian Pines dataset (INDI) recorded by the AVIRIS airborne hyperspectral imager over North-west Indiana, USA, is made available by Purdue University and comprises 145x145 pixels at 20 m spatial resolution and 224 spectral reflectance bands from 400-2500 nm (Baumgardner, Biehl et al. 2015). Removal of water absorption bands by the provider reduced these to 200 wavebands, and then reflectance of each pixel was scaled between 0 and 1. Fifty pixels were randomly selected as training samples except for three classes with fewer than 50 total samples, for which 15 samples were used for training (Table 3).

The New Zealand (NZ) dataset used in this study is a sub-sample of hyperspectral spectra for 22 species taken from a dataset of 39 native New Zealand plant spectra collected from four different sites around the North Island of New Zealand and made available on the SPECCHIO database (Hueni, Chisholm et al. 2020). These spectra were acquired with an ASD FieldSpecPro spectroradiometer at 1 nm sampling intervals between 350 – 2500 nm. Following acquisition from the SPECCHIO database, spectra were resampled to 3 nm and noisy bands associated with atmospheric water absorption were removed (1326 – 1464, 1767 – 2004, 2337 – 2500) resulting in 540 bands per spectra. Eighty percent of samples per class were used for training the GAN and 20 % held aside to evaluate classifier performance (Table 4).

Table 3. Land cover classes, training and evaluation sample numbers for Indian Pines dataset.

Class ID	Class name	Training samples	Evaluation samples
INDI1	Alfalfa	15	31
INDI2	Corn-no-till	50	1378
INDI3	Corn-min-till	50	780
INDI4	Corn	50	187
INDI5	Grass-pasture	50	433
INDI6	Grass-trees	50	680
INDI7	Grass-pasture-mowed	15	13
INDI8	Hay-windrowed	50	428
INDI9	Oats	15	5
INDI10	Soybean-no-till	50	922
INDI11	Soybean-min-till	50	2405
INDI12	Soybean-clean	50	543
INDI13	Wheat	50	155
INDI14	Woods	50	1215
INDI15	Buildings-Grass-Trees-Drives	50	336
INDI16	Stone-Steel-Towers	50	43

Table 4. Plant species classes, training and evaluation sample numbers for New Zealand dataset.

Class ID	Common name	Botanical name	Training samples	Evaluation samples
NZ0	Manuka	*Leptospermum scoparium*	58	14
NZ1	Pohutukawa	*Metrosideros excelsa*	32	8
NZ2	Koromiko	*Hebe stricta*	42	10
NZ3	Lemonwood	*Pittosporum eugenioides*	46	12
NZ4	Kawakawa	*Macropiper excelsum*	34	9
NZ5	Whiteywood	*Melicytus ramiflorus*	48	12
NZ6	Totara	*Podocarpus totara*	34	8
NZ7	New Zealand Flax	*Phormium tenax*	36	9
NZ8	Akiraho	*Olearia paniculata*	8	2
NZ9	Rata	*Metrosideros robusta*	9	2
NZ10	Ngaio	*Myoporum laetum*	38	10
NZ11	Mapou	*Myrsine australis*	36	9
NZ12	Cabbage tree	*Cordyline australis*	32	8
NZ13	Karaka	*Corynocarpus laevigatus*	34	9
NZ14	Kauri	*Agathis australis*	15	3
NZ15	Silver fern	*Cyathea dealbata*	28	7
NZ16	Tangle fern	*Gleichenia dicarpa var. alpina*	14	4
NZ17	Black tree fern	*Cyathea medullaris*	18	4
NZ18	Pigeonwood	*Hedycarya arborea*	18	5
NZ19	Rangiora	*Brachyglottis repanda*	12	3
NZ20	Karamu	*Coprosma robusta*	13	3
NZ21	Red Pine	*Dacrydium cupressinum*	16	4

3.5. Results & Discussion

3.5.1 Mean and standard deviation of training and synthetic spectra

In order to visualize similarities between synthetic and real spectra the mean and standard deviation for each class are shown for the real, evaluation, and synthetic datasets. All low frequency spectral features, as well as mean and standard deviations appear to be reproduced with high accuracy by the GAN. At finer scales of 3-5 wavebands noise is present, most notably throughout the near infra-red (NIR) plateau (Figure 9). Smoothing of synthesized data by a number of methods resulted in either no improvement or decreased performance in a number of tests; for this reason, no pre-processing was performed on synthesized samples. Due to the high frequency and random nature of the noise, once mean and STD statistics are calculated the spectra appear smooth.

Class 0 is one of the NZ classes with the largest number of samples, resulting in its mean and standard deviation being similar between its real, evaluation, and synthetic subsets. However, this is not the case for all classes, with NZ-9 showing the mean and standard deviation of the randomly selected evaluation samples being vastly different to those of real and synthetic spectra (Figure 10). The same is seen amongst INDI classes, with class 2 matching across all 3 data subsets, and class 4 with only 40 samples showing substantial difference between evaluation and real samples, especially in the visible wavebands (Figure 11). Although some classes may struggle to represent the evaluation dataset due to the initial random splitting of the datasets, in general mean and standard deviation of the synthetic samples very closely match the real training data.

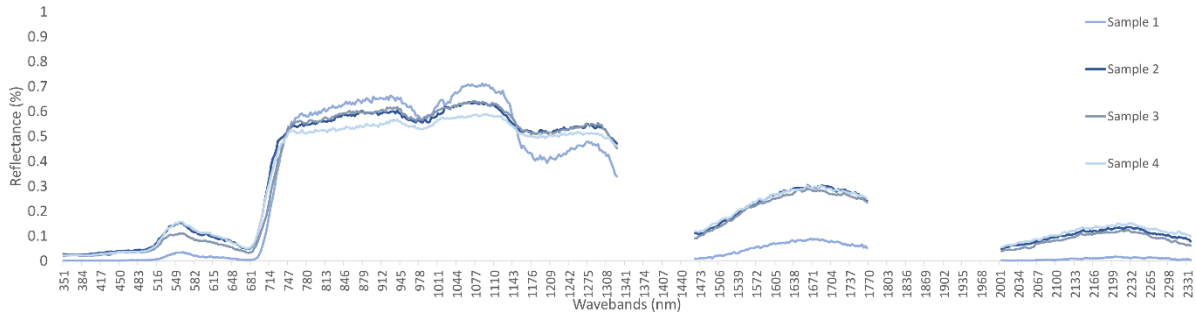


Figure 9. Synthetic spectra of NZ class 0, 350-2400nm at 3nm bandwidths

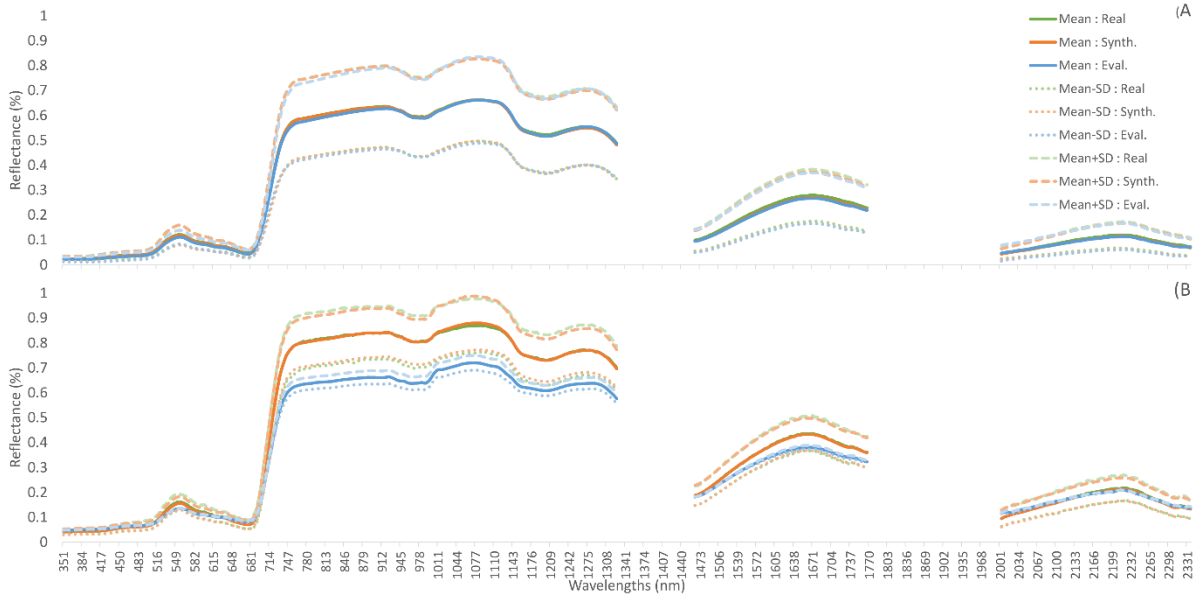


Figure 10. Mean and +/- 1 STD for training (real), synthetic, and evaluation (real) datasets. (A NZ class 0; Manuka (*L. scoparium*), (B NZ class 9; Rata (*M. robusta*).

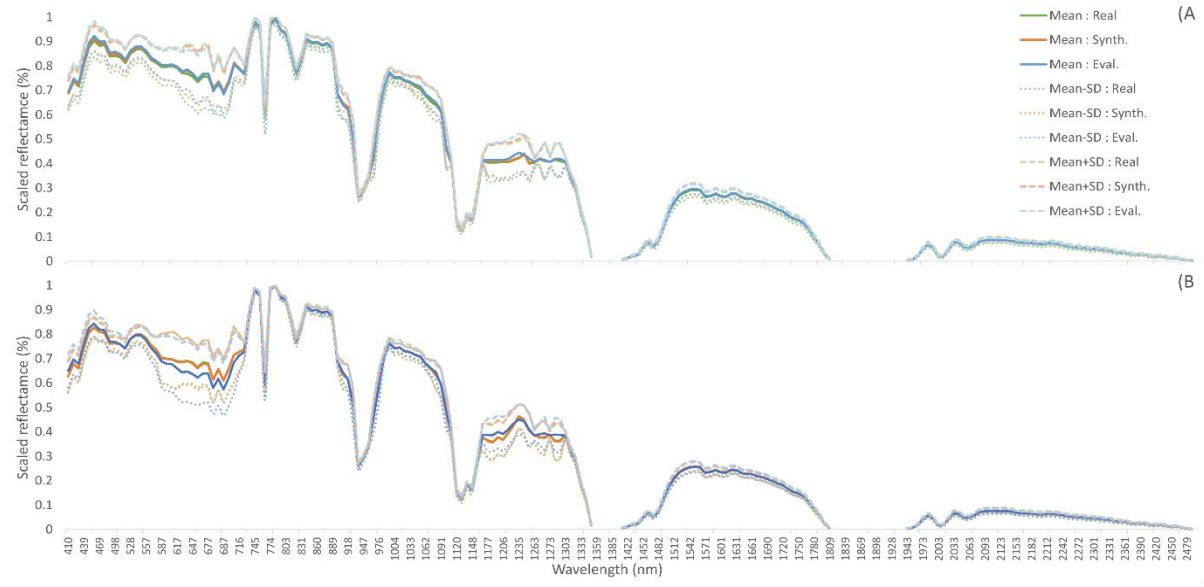


Figure 11. Mean and +/- 1 STD for training (real), synthetic, and evaluation (real) datasets. (A INDI class 2; Corn-notill, (B INDI class 4; Corn.

3.5.2 Generation & distribution of spectra

Here we demonstrate the ability of the GAN to reproduce realistic spectral shapes and to capture the statistical distribution of the class populations. Three dimensionality reduction methods — PCA, t-SNE, and UMAP — were applied to both the real and synthetic datasets of INDI and NZ spectra to reduce their 200 and 540 wavebands (respectively) down to a plottable 2D space (Figure 12, 13). Upon visual inspection the class clusters formed by the augmented data across all reduction methods mimic the distribution of those of the real data. Additionally, due to its small sample sizes the structure of clusters for the real NZ data is sparse and unclear, though is emphasised by the large number of synthetic samples.

Such strong replication of the 2D representation of the classes is a good indication of the generative model's ability to learn distributions. Even when the models are trained separately for each class the relationship between classes is maintained. However, the increased sample number in the synthetic datasets do in some cases extend beyond the bounds of the real samples. Whilst some may represent potential outliers the majority are artefacts of increased sample sizes. This is most evident in the UMAP representation where a parameter that defines minimum distance between samples can be set, a larger value of which results in increased spread of samples in the 2D representation (McInnes, Healy et al. 2018). This is most notable in the INDI dataset, with classes 1, 7, and 8 extending more broadly than the real dataset (Figure 12F).

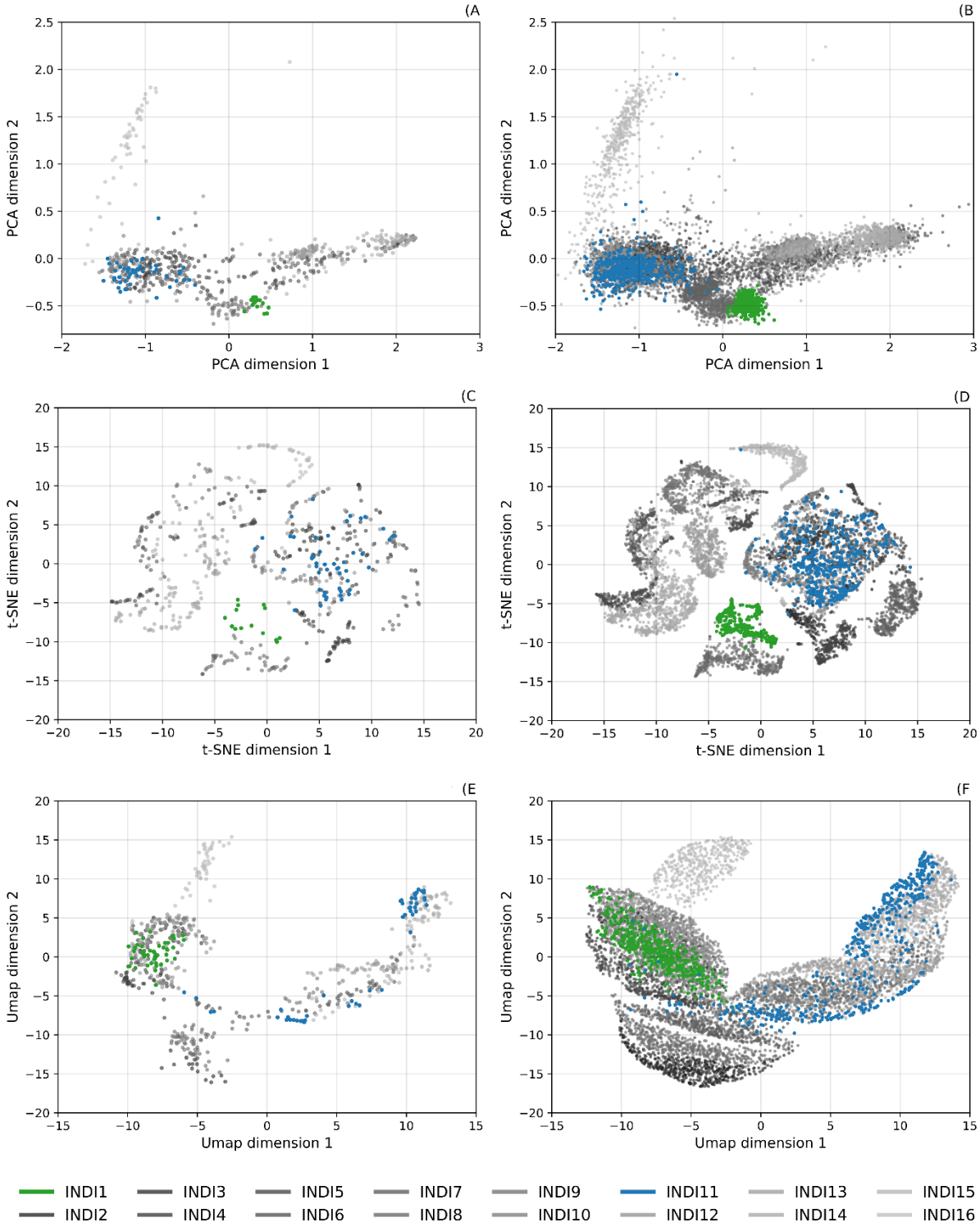


Figure 12. Dimensional reduced representations of INDI real and synthetic datasets; highlighted classes INDI1 - Alfalfa (green), INDI11 – Soybean-min-till (blue). A) Real dataset; PCA reduction, B) Synthetic dataset; PCA reduction, C) Real dataset; t-SNE reduction, D) Synthetic dataset; t-SNE reduction, E) Real dataset; UMAP reduction, F) Synthetic dataset; UMAP reduction.

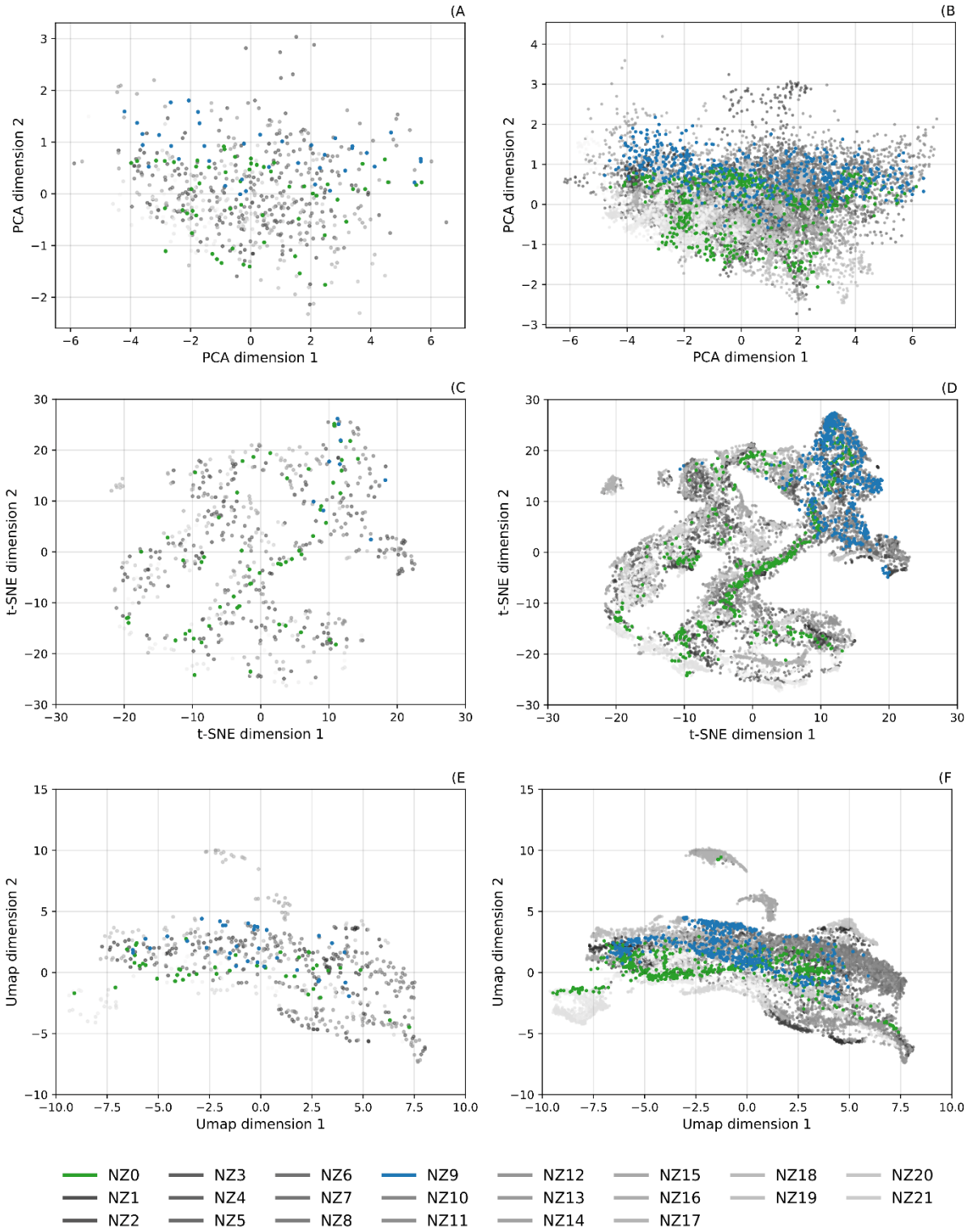


Figure 13. Dimensional reduced representations of NZ real and synthetic datasets; + highlighted classes NZ0 – Manuka (L. scoparium) (green), NZ9 – Rata (M. robusta) (blue). A) Real dataset; PCA reduction, B) Synthetic dataset; PCA reduction, C) Real dataset; t-SNE reduction, D) Synthetic dataset; t-SNE reduction, E) Real dataset; UMAP reduction, F) Synthetic dataset; UMAP reduction.

3.5.3 Training classification ability

In order to further examine the similarity of synthetic spectra to the real training data, three classifiers were trained (SVM, RF, NN), with four permutations of each (real – real, real – synthetic, synthetic – synthetic, and synthetic - real) (Table 5). With few exceptions the neural network classifier outperformed the others, with SVM being the second most accurate, followed by RF. The INDI dataset recorded highest accuracy for real - real test with RF and NN classifiers at 74.76 % and 84.13 % respectively, although the highest accuracy for the SVM classifier occurred during the synthetic- real test with 81.42 %. Comparing the four combinations of real and synthetic, real-real had the highest accuracy for four experiments, with INDI synthetic-real with the SVM, and NZ real-synthetic with the RF classifier being the only exceptions.

Table 5. Classification accuracies for classifiers trained on real or synthesised spectral data and evaluated on either real or synthesised data for both Indian Pines and New Zealand datasets based on real class sample sizes.

INDI	SVM	RF	NN
Real - Real	73.48	**74.76**	**84.13**
Real - Synthetic	77.04	66.68	76.50
Synthetic - Synthetic	79.48	69.38	80.91
Synthetic - Real	**81.42**	70.51	81.66
NZ			
Real - Real	**79.86**	47.65	**95.76**
Real - Synthetic	78.73	**60.23**	80.54
Synthetic - Synthetic	74.20	51.33	91.82
Synthetic - Real	78.19	54.76	81.13

To further evaluate the synthetic spectra, synthetic-synthetic and synthetic-real experiments were performed with the number of synthesized training samples increasing from 10 to 490 samples by increments of 10 samples per class (Figure 14). Synthetic-synthetic accuracy improves with more samples: this too is to be expected as this simply adds more training samples from the same distribution. Most importantly synthetic-real accuracy, though often slightly lagging synthetic-synthetic improves in the same manner, indicating that the synthetic samples are a good representation of the true distribution and that increasing their number for training a classier is an effective method of data augmentation. The main exception to this is the NN NZ classier, where synthetic-synthetic quickly reaches ~100 % accuracy while synthetic-real

maintains ~80 % before slowly decreasing in accuracy as more samples are added. This could indicate the NN classifier focuses on different features than the other classifiers, potentially being more affected by the small-scale noise apparent in the NZ generated samples as the noise isn't as apparent in the INDI data and the INDI NN classifier doesn't show such a discrepancy between synthetic-synthetic and synthetic-real.

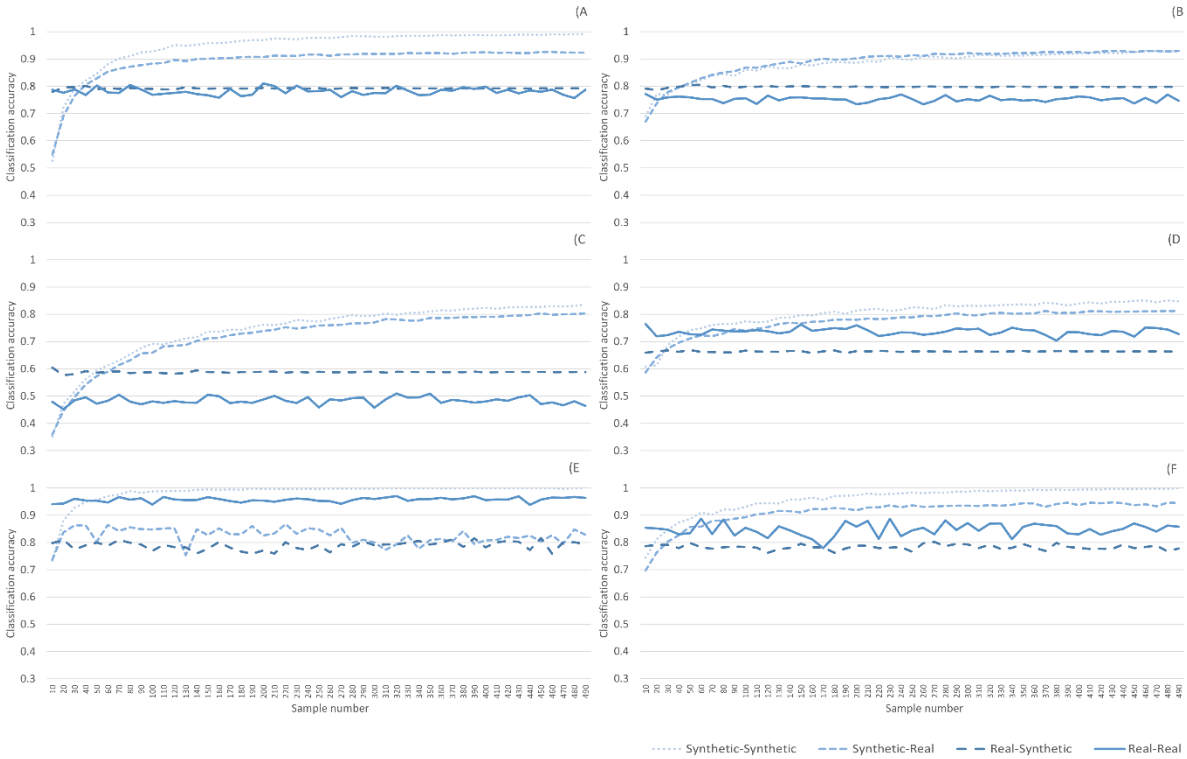


Figure 14. Classification accuracies for classifiers trained on real or synthesised spectral data and evaluated on either real or synthesised data for both Indian Pines and New Zealand datasets ranging from 10 - 490 samples per class. (A New Zealand dataset; SVM classifier, (B Indian Pines dataset; SVM classifier, (C New Zealand dataset; RF classifier, (D Indian Pines dataset; RF classifier, (E New Zealand dataset; NN classifier, (F Indian Pines dataset; NN classifier.

3.5.4 Data augmentation

In order to test the viability of the synthetic data for data augmentation the same three classifiers were trained with either real, synthetic or both combined into an augmented dataset and tested against an evaluation dataset (Table 6). All classifiers had higher accuracy when trained on the real dataset compared to synthetic, though the highest accuracy overall was with the augmented dataset. For the INDI data this increase was minor, being < 1 % for all classifiers. A far more significant improvement was seen for the NZ data with increases of 3.54 % (to 86.55 %), 0.53 % (to 50.80 %), and 3.73 % (to 85.14 %) for SVM, RF, and NN respectively.

Table 6. Classification accuracies for classifiers trained on real, synthesised, or augmented spectral data and evaluated on an evaluation dataset for both Indian Pines and New Zealand datasets based on real class sample sizes.

INDI	SVM	RF	NN
Real – Evaluation	70.40	66.40	62.06
Synthetic - Evaluation	62.82	57.73	51.30
Augmented - Evaluation	**70.56**	**66.91**	**62.76**
NZ			
Real - Evaluation	83.01	50.27	81.41
Synthetic - Evaluation	63.02	36.69	68.60
Augmented - Evaluation	**86.55**	**50.80**	**85.14**

Of course, however, the number of synthetic samples does not have to be limited in such a manner. As with previous experiments the number of synthetic samples started at 10 and incremented by 10 to a total of 490, demonstrating the potential of this data augmentation method. Dramatic increases in accuracy were seen for the synthetic dataset, with the smallest increase being 5.13 % for INDI-SVM occurring at 490 samples, the largest being 20.47 % for NZ-RF at 420 samples. These increases brought the synthetic dataset very close to the accuracy of the real samples or even above in the cases of INDI-NN, NZ-RF, and NZ-NN. Increases in accuracy were also seen in the augmented dataset, though not as dramatic as those for the synthetic dataset. Improvements in accuracy ranged from 0.16 % for INDI-SVM at 10 synthetic samples to 9.45 % for NZ-RF at 280 synthetic samples. These improvements raise the highest accuracy for the INDI dataset from 70.40 % to 70.56 %, resulting in an increase of 0.16 % over the highest achieved by just the real data. A larger increase was seen in the NZ dataset with the previous highest accuracy raising from 86.55 % to 90.01 %, an increase of 3.45 % from the previous augmented classification with restricted sample size, and a 7% increase over the real dataset alone (Table 7).

Table 7. Classification accuracies for classifiers trained on real, synthesised, or augmented spectral data and evaluated on an evaluation dataset for both Indian Pines and New Zealand datasets with sample sizes ranging from 10 to 490 per class for synthetic and augmented while real contained all real samples.

INDI	SVM / Sample size	RF / Sample size	NN / Sample size
Real – Evaluation	70.40 / All	66.40 / All	62.06 / All
Synthetic - Evaluation	67.95 / 490	65.59 / 490	65.05 / 50
Augmented - Evaluation	**70.56 / 10**	**68.25 / 140**	**69.77 / 320**
NZ			
Real - Evaluation	83.01 / All	50.27 / All	81.41 / All
Synthetic - Evaluation	81.35/ 490	57.16 / 420	87.78 /450
Augmented - Evaluation	**90.01/ 120**	**60.25 / 280**	**89.25 / 120**

3.5.5 Classification power of a synthetic sample

Ordering the synthetic samples by their C value before iteratively adding a single sample at a time from each class to the training dataset of an SVM classifier shows the differing classification power of the synthetic samples from lower to upper bounds of C and vice versa (Figure 15).

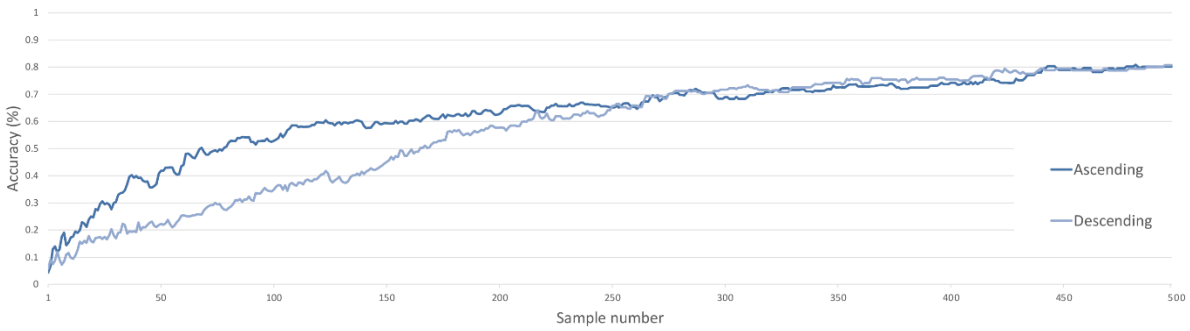


Figure 15. Classification accuracy of a SVM classifier for C metric ascending and descending ordered synthetic datasets incremented by single samples.

When in ascending order from lower to upper bounds classification accuracy increases dramatically, reaching ~60 % accuracy with ~100 samples, while 200 samples were required for similar accuracy in descending order. At approximately half the number of samples accuracies converge, then increase at the same rate before reaching 80 % accuracy at 500 samples. These classification accuracies (Figure 15) provide the first insight into increased discriminatory power associated with synthetic samples that occur at distance to real samples. Although not encountered here, a maximum limit to

this distance would be present with synthetic samples needing to remain within the bounds of their respective class distributions.

A similar though reduced trend can be seen when the ordered synthetic samples are used to augment the real dataset. Both ascending and descending datasets improve classification over that of the real dataset when samples are iteratively added to the classifiers training dataset (Figure 16). Despite descending ordered samples outperforming ascending at times, on average ascending achieved ~1.5 % higher accuracy across the classifications with an average value of 79.72 % to 78.24 % of descending samples.

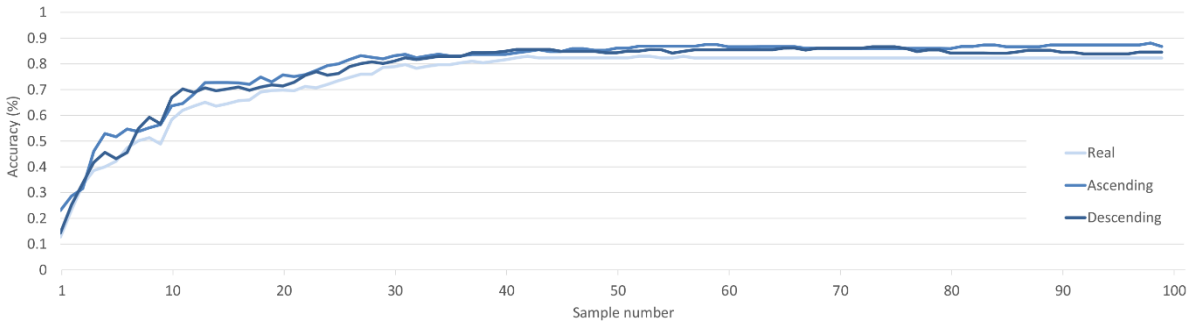


Figure 16. Classification accuracy of a SVM classifier for C metric ascending and descending augmented datasets with randomly ordered real dataset incremented by single samples.

This artificial selection of synthetic data points distant or close to the real data influences sample distribution used to train the classifiers. As one might expect the ordered data points come from the edges or sparse regions of the real data distribution, dramatically shifting the mean and standard deviation of the ordered datasets (Figure 17).

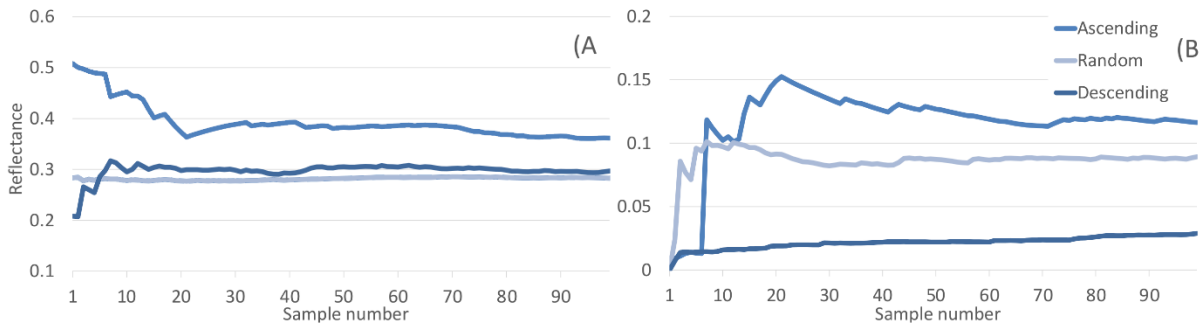


***Figure 17**. A) Mean and B) STD of C metric ascending, descending, and randomly ordered synthetic datasets incremented by single samples.*

The inclusion of synthetic data points selected at random provides a baseline for comparison with the ordered datasets. Once the number of samples increases beyond a few points the means for descending and random converge and stay steady throughout. Mean values for ascending start significantly higher, though initially begin to converge towards the other datasets before plateauing at a higher level. Whilst being averaged across all classes and all wavebands of spectra, the mean reflectance for the ascending data is consistently higher. Standard deviation of the descending dataset is consistently low, only slightly increasing as samples are added. This is in stark difference to the STD of the ascending dataset being ~5 – 6 x higher across all n samples. The mean of the randomly selected dataset occurs between the means of the two ordered, though closer to the ascending mean, indicating the samples that make up the descending dataset are highly conserved.

To further illustrate the relationship of the ordered datasets and the real distribution, a PCA of one of the classes is shown (Figure 18). As the mean and STD indicated the descending samples are tightly grouped near the mean and densest area of the real data distribution, with the ascending samples generally occurring along the border of the real distribution. Whilst ascending selects for samples with low C and greater distance from real samples it is important to note that these synthetic samples still appear to conform to the natural shape of the real distribution, a further indication the generative model is performing well.

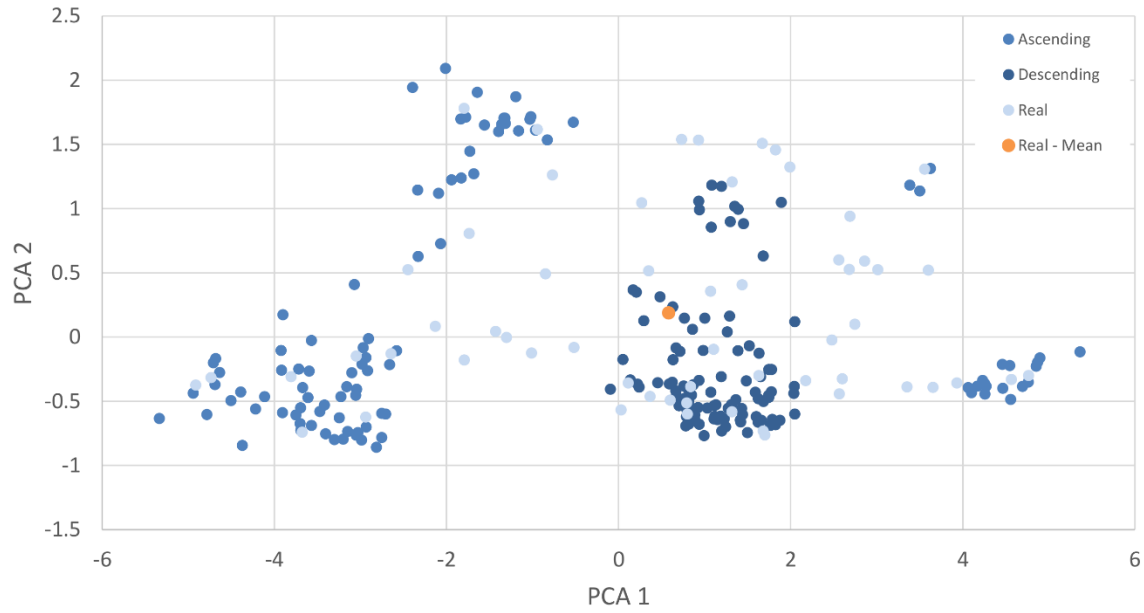


Figure 18. PCA of NZ class 0; Manuka (*L. scoparium*) real samples, with first 100 samples of the ascending and descending C ordered synthetic datasets.

3.6. Conclusions

In this paper we have successfully demonstrated the ability to train a generative machine learning model for synthesis of hyperspectral vegetation spectra. Evaluation of the synthetic spectra shows that they respect many of the statistical properties of the real spectra, conforming well to the sampled distributions of all real classes. Further to this we have shown that the synthetic spectra generated by our models are suitable for data augmentation of a classification models training dataset. Addition of synthetic samples to the real training samples of a classifier produced increased overall classification accuracy under almost all circumstances examined. Of the two datasets, the New Zealand vegetation showed a maximum increase of 7.0 % in classification accuracy with Indian Pines demonstrating a more modest improvement of 0.16 %. Selection of synthetic samples from sparse or outlying regions of the feature space of real spectral classes demonstrated increased discriminatory power over those from more central portions of the distributions. We believe further work regarding this could see targeted generation to maximize the information content of a synthetic sample that would result in improved classification accuracy and generalizability with a smaller augmented dataset. The use of these synthesized spectra to augment real spectral datasets allows for the training of classifiers that benefit from large sample numbers

without a researcher needing to collect additional labelled spectra from the field. This is of increasing significance as modern machine and deep learning algorithms tend to require larger datasets.

3.7. References

Arjovsky, M., S. Chintala and L. Bottou (2017). "Wasserstein gan." *arXiv preprint arXiv:1701.07875*.

Audebert, N., B. Le Saux and S. Lefèvre (2018). Generative adversarial networks for realistic synthesis of hyperspectral samples. IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, IEEE.

Baumgardner, M. F., L. L. Biehl and D. A. Landgrebe (2015). 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3.

Bellemare, M. G., I. Danihelka, W. Dabney, S. Mohamed, B. Lakshminarayanan, S. Hoyer and R. Munos (2017). "The cramer distance as a solution to biased wasserstein gradients." *arXiv preprint arXiv:1705.10743*.

Bissoto, A., F. Perez, E. Valle and S. Avila (2018). "Skin lesion synthesis with generative adversarial networks." In OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, Springer**:** 294-302.

Donahue, C., J. McAuley and M. Puckette (2018). "Adversarial audio synthesis." *arXiv preprint arXiv:1802.04208*.

Esteban, C., S. L. Hyland and G. Rätsch (2017). "Real-valued (medical) time series generation with recurrent conditional gans." *arXiv preprint arXiv:1706.02633*.

Feng, J.; Yu, H.; Wang, L.; Cao, X.; Zhang, X.; Jiao, L. (2019) "Classification of Hyperspectral Images Based on Multiclass Spatial–Spectral Generative Adversarial Networks." *IEEE Trans. Geosci. Remote Sens. 57*, 5329–5343.

Goodenough, A. A. and S. D. Brown (2017). "DIRSIG5: next-generation remote sensing data and image simulation framework." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10(11): 4818-4833.

Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio (2014). "Generative adversarial nets." *NIPS*

Gui, J., Z. Sun, Y. Wen, D. Tao and J. Ye (2020). "A review on generative adversarial networks: Algorithms, theory, and applications." *arXiv preprint arXiv:2001.06937*.

Gulrajani, I., F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville (2017). "Improved training of wasserstein gans." *arXiv preprint arXiv:1704.00028***.**

Harada, S., H. Hayashi and S. Uchida (2019). "Biosignal Generation and Latent Variable Analysis With Recurrent Generative Adversarial Networks." *IEEE Access* PP: 1-1.

Hennessy, A., K. Clarke and M. Lewis (2020). "Hyperspectral Classification of Plants: A Review of Waveband Selection Generalisability." *Remote Sensing* 12(1): 113.

Hueni, A., L. A. Chisholm, C. Ong, T. J. Malthus, M. Wyatt, S. A. Trim, M. E. Schaepman and M. Thankappan (2020). "The SPECCHIO Spectral Information System." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 5789-5799.

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

JPL/NASA. ECOSTRESS. (2021), Available online: https://ecostress.jpl.nasa.gov/ (accessed on 28/05/2021).

LeCun, Y. (2016). What are some recent and potentially upcoming breakthroughs in deep learning. Quora. https://www.quora.com/What-are-some-recent-and-potentially-upcoming-breakthroughs-in-deep-learning.

McInnes, L., J. Healy and J. Melville (2018). "Umap: Uniform manifold approximation and projection for dimension reduction." *arXiv preprint arXiv:1802.03426*.

Mountrakis, G. and B. Xi (2013). "Assessing reference dataset representativeness through confidence metrics based on information density." *ISPRS Journal of Photogrammetry and Remote Sensing* 78: 129-147.

Nalepa, J., M. Myller, M. Kawulok, B. Smolka, N. Kehtarnavaz and M. F. Carlsohn (2019). “On data augmentation for segmenting hyperspectral images.” Real-Time Image Processing and Deep Learning 2019.

Slavkovikj, V., S. Verstockt, W. De Neve, S. Van Hoecke and R. Van de Walle (2015). Hyperspectral Image Classification with Convolutional Neural Networks. Proceedings of the 23rd ACM International Conference on Multimedia - MM '15**:** 1159-1162.

Song, J. (2017). cramer-gan. https://github.com/jiamings/cramer-gan, GitHub.

Taylor, L. and G. Nitschke (2017). "Improving deep learning using generic data augmentation." *arXiv preprint arXiv:1708.06020*.

Wang, G., W. Kang, Q. Wu, Z. Wang and J. Gao (2018). Generative adversarial network (GAN) based data augmentation for palmprint recognition. 2018 Digital Image Computing: Techniques and Applications (DICTA), IEEE.

Wang, K. (2019). "Synthetic data generation and adaptation for object detection in smart vending machines." *arXiv preprint arXiv:1904.12294*.

Wen, Q., L. Sun, X. Song, J. Gao, X. Wang and H. Xu (2020). "Time Series Data Augmentation for Deep Learning: A Survey." *arXiv preprint arXiv:2002.12478*.

Xu, Y.; Du, B.; Zhang, L. (2018) “Can We Generate Good Samples for Hyperspectral Classification? — A Generative Adversarial Network Based Method.” In Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium; Institute of Electrical and Electronics Engineers (IEEE), 5752–5755.

Zhan, Y.; Hu, D.; Wang, Y.; Yu, X. (2017) “Semisupervised Hyperspectral Image Classification Based on Generative Adversarial Networks.” *IEEE Geosci. Remote. Sens. Lett. 15*, 212–216.

Zhu, L., Y. Chen, P. Ghamisi and J. A. Benediktsson (2018). "Generative Adversarial Networks for Hyperspectral Image Classification." *IEEE Transactions on Geoscience and Remote Sensing* 56(9): 5046-5063.

Intentionally left blank

CHAPTER FOUR: Generative adversarial network-based style transfer for data augmentation of hyperspectral vegetation data

This chapter is submitted for publication as:

Hennessy, A.; Clarke, K.; Lewis, M. (*Submitted*) "Generative adversarial network-based style transfer for data augmentation of hyperspectral vegetation data."

Statement of Authorship

Title of Paper	Generative adversarial network-based style transfer for data augmentation of hyperspectral vegetation data	
Publication Status	☐ Published	☐ Accepted for Publication
	☒ Submitted for Publication	☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Hennessy, A.; Clarke, K.; Lewis, M. Generative adversarial network-based style transfer for data augmentation of hyperspectral vegetation data. (unpublished)	

Principal Author

Name of Principal Author (Candidate)	Andrew Hennessy	
Contribution to the paper	Conceptualization, Methodology, data curation, formal analysis, writing – original draft preparation, writing – review and editing.	
Overall percentage (%)	70%	
Certification	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.	
Signature		Date: 13/07/2021

Co-author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Kenneth D. Clarke	
Contribution to the paper	Methodology, formal analysis, writing – review and editing	
Signature		Date: 13/07/2021

Name of Co-Author	Megan M. Lewis	
Contribution to the paper	Writing – review and editing	
Signature		Date: 12/07/2021

Abstract

The significant issue of intra-class variance and inter-class overlap has severely reduced the generalisability of hyperspectral vegetation classification models, resulting in many studies being singular and ad-hoc, lacking in spatial or temporal transferability. In order to rectify this, we apply advances in generative deep learning models to produce realistic synthetic hyperspectral vegetation data, whilst transforming them based upon the learned features (styles) of the other classes in the study. Specifically, a generative adversarial network (GAN) is trained on a vegetation hyperspectral dataset, demonstrating the ability to manipulate their intra/inter-class relationships in a desirable manner.

Transformations were performed on both the training and evaluation subsets of a hyperspectral vegetation dataset producing *n* transformed replicates of every sample where *n* is the number of classes in the study. Training classification models and then evaluating subsets with the same style transformations produced accuracies significantly higher than those of the original dataset. An accuracy of 97.38 % was achieved for a single transformed dataset, an increase of 11.95 % over the original dataset classification accuracy of 85.43 %. This significant increase in accuracy was then surpassed following the ensembling of all transformed subsets classification results, reaching an accuracy of 98.62 %. Evaluation of the transformed synthetic spectra shows the samples to be tightly clustered by class, indicating the successful reduction of intra-class variance as well the reduction in inter-class overlap.

4.1. Introduction

Hyperspectral (HS) imagery and field recorded spectra contains information detailing the structure and biochemistry from the surface and internals of plants leaves, as well as overall architecture and plant structure from branches and trunks. Extracting this information from the HS measurements can offer insights into plant growth stage, general health, and even taxonomy. Identification of plant types is a common goal in remote sensing, from locating invasive species, forestry monitoring, agriculture and biodiversity monitoring.

Supervised classification of plants from HS data relies upon the members of a taxonomic or functional group having similar spectral features to each other whilst having features distinct from members of other groups within the data. No matter the taxonomic or phenological resolution of the classes being classified, accuracy is dependent upon minimising intra-class variability while maximizing inter-class.

Driven by environmental conditions such as nutrient, sunlight and water availability, as well as growth-stage and seasonality effects, intra-class variability within leaf spectra of a single plant species can be substantial (Castro-Esau, Sánchez-Azofeifa et al. 2006). For this reason, classification models are often found to have limited generalisability and transferability across spatial and temporal distances, with collection of further labelled reference samples and training of new classification models required for wider application. This issue plays into the “one place, one time” problem identified by Woodcock (2002), where studies are often spatially and temporally constrained limiting their insight and applicability to other related studies.

With the launch of PRISMA (Lopinto and Ananasso 2020), as well as further planned HS satellites in the near future and technological advancements and reductions in cost making drone and airborne HS measurements increasingly commonplace more and more of the environment is being recorded as HS data. In order to convert these measurements into actionable knowledge the aforementioned limitations of generalisability and transferability need to be overcome. Having previously addressed this issue with generative models and data augmentation with some degree of success (Hennessy, Clarke et al. 2021), as has others (Nicolas Audebert 2018), we now

investigate a method analogous to neural style transfer (NST). Conventionally, data augmentation aims to expand an available training dataset by synthesizing additional novel views of the original data, whereas here we explore the use of generative models to synthesize transformations of the original data in a way that moves the intra-class/inter-class balance in our favour. Rather than expanding the training dataset to better encapsulate the evaluation data, here we aim to demonstrate that such NST style transformations can force the evaluation data into the training data's feature space respective of classes. Contraction of a classes distribution within feature space in such a manner would decrease intra-class variance, however, the ultimate success in improving classification accuracy and generalisability will be determined by the transformations effects on inter-class distances.

Originally NST was performed on imagery with a convolutional neural network specifically designed for the task of separating style and content of images before recombining them in novel ways (Gatys, Ecker et al. 2016). Now a fast developing field of its own with academic, artistic and industrial implementations (Jing, Yang et al. 2019) we borrow inspiration from NST to synthesis unique transformations of HS vegetation data. We say analogous to NST as the method presented here does not explicitly learn styles or contents of the data. Rather we sample from either interpolated regions of the learned latent space or from latent vectors representing regions outside of the training data's feature space, both of which result in replicated samples that exhibit stylized features of other classes. For what we believe to be the first time reported in the literature we go on to use these stylized replicated spectra in an ensemble classification model to produce dramatic increases in classification accuracy.

4.2. Experimental design

Here we implement a further refined version of the CramérGAN architecture utilized by Hennessy, Clarke et al. (2021). While that approach generated realistic vegetation spectra with minimal noise and no signs of mode collapse, modifications were required to improve learning stability and interpretability of the model's latent space.

The most significant change in this study from the previous approach lies with the latent noise vector. Previously this 100 length vector was generated from a normal distribution before being sent to the generator (Hennessy, Clarke et al. 2021). Although

still generated from a normal distribution, in our new approach the latent vector consists of 3 smaller vectors with lengths 64, 16, and 16 which were concatenated together to a total length of 96. The 64 length vector is a simple normally distributed noise vector, the first 16 length vector was generated as a 4 length normally distributed noise vector before being reshaped to a 2x2 matrix that was 2 x convolved to shape 4x4 then flattened. The second 16 length vector was generated as a single random value from a normal distribution before being duplicated to length 16 (Figure 19). This latent structure was based on Alharbi and Wonka (2020) to aid in disentangling the learned latent space and making the latent space to feature space back to latent space mapping more coherent. They demonstrated the injection of noise in this grid-based structure provided greater generative control over the model without requiring labelled data or modifications to the network's architecture.

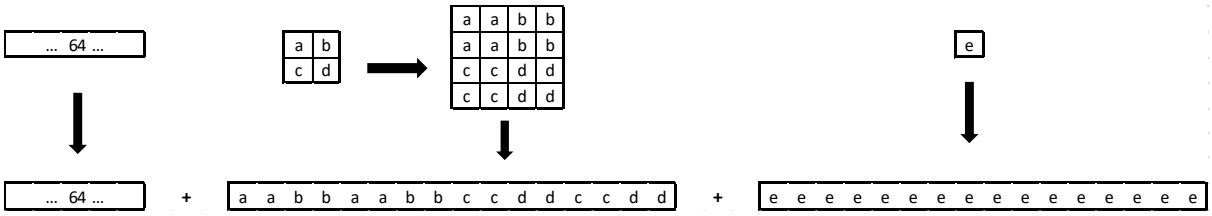


Figure 19. Construction of a 96 length latent vector from 3 smaller vectors sampled from a normal distribution with shapes 64,4,1 before being expanded to 64,16,16 and concatenated together.

Generative models were trained for 100,000 epochs, at a ratio of 5:1 (5 training iterations of D for every 1 of G) using the RSMP optimizer at a learning rate of 0.0001, with beta1 = 0.5 and beta2 = 0.9. The G consists of two fully connected dense layers followed by two convolution layers, all using the ReLu activation function save for the final convolution layer using Sigmoid activation. The final layer of G reshaped the output to be a 2D array with shape (batch size * number of bands) (Table 8). A simpler architecture was used for the D, consisting of 3 fully connected dense layers, all layers of D used Leaky ReLu activation except the final layer which used a linear function (Table 9).

Table 8. Layer architecture of the GANs Generator.

Layer type / parameters	Shape	Activation
Conv1D	(100,100)	ReLU
Conv1D	(100,50)	ReLU
Conv1D	(100,10)	ReLU
Flatten		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	4096	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	22	Softmax

Table 9. Layer architecture of the GANs Discriminator.

Layer type / parameters	Shape	Activation
Dense	1024	Leaky ReLU (alpha = 0.2)
Dense	1024	Leaky ReLU (alpha = 0.2)
Dense	256	Linear

Generative models were trained on either all classes or individual models for each class for all samples in the training data subset (Table 11). Following training, synthetic datasets were generated using randomly sampled latent vectors with 10,000 samples for the combined all class model and 5,000 samples from each individual class model.

Each synthetic sample paired with its generating latent vector was used to train a multi-regression analysis (MRA) with the synthetic spectral sample as the independent axis and its latent vector as the dependent. In order to test prediction accuracy and speed of analysis six variants of the MRA were performed for each sample with varying numbers of features being passed as independent variables. Number of features were all 540 wavebands, every fifth for 108, tenth for 54, twentieth for 27, twenty seventh

for 20, and fifty fourth for 10 wavebands. Predicted latent vectors were generated for the training and evaluation subsets which were in turn passed as input to the generator in order to generate a synthetic replicate of each of the training and evaluation samples. Accuracy of these generated replicas were assessed by calculating the difference between the replica and its real counterpart for the total Euclidean distance between the spectra and a number of hyperspectral indices (Table 10).

Table 10. Hyperspectral indices used in replication comparison.

Name	Equation	Reference
Modified Simple Ratio 705/445	750nm − 445 nm / 705 nm − 445 nm	Sims and Gamon (2002)
Anthocyanin reflectance index	1 / 550 nm – 1 / 700 nm	Gitelson, Merzlyak et al. (2001)
Cellulose Absorption Index 2	0,5(2020 nm + 2220 nm) – 2100 nm	Nagler, Daughtry et al. (2000)
Difference 1725/970 Difference LAI	1725 nm − 970 nm	Le Maire, François et al. (2008)
Disease water stress index	802 nm + 547 nm / 1657 nm + 682 nm	Galvao, Formaggio et al. (2005)
Normalized Difference 1094/1205 Leaf water VI 2	1094 nm − 1205 nm / 1094 nm + 1205 nm	Galvao, Formaggio et al. (2005)

Upon selection of the best performing MRA hyperparameters the final regressed replicates of the real spectra were generated. Either all samples were generated from the single all class trained model (Figure 21), or every sample was generated from each individual class-trained model, resulting in 22 alternate versions for each class. To assess the classification potential of all versions three classifiers were trained on various combinations of real and replicated subsets: these were a support vector machine (SVM), random forest (RF), and neural network (NN) (Figure 20).

Selection of a single subset to represent a class from the 22 alternative subsets per class was performed. The method for this selection was based on the idea that the replicate produced from its correct class GAN model would be most similar to the original sample that was used in the regression to produce its latent vector. Similarities were calculated by distance in feature space between the real and replicate samples. Post processing of the spectra was performed before the distance calculations to help emphasise differences between samples. Distances were calculated on either the raw spectra, albedo normalised spectra based upon (Pu, Bell et al. 2012), or 1st or 2nd order derivatives created with the Savitzky-Golay smoothing algorithm.

While the single class GAN models inherently produce style transferred replicates, the single all-class GAN model requires a different approach. Interpolation between latent vectors to produce smooth transitions between generated features is commonplace in the generative machine learning community for generating novel views of the learned data and exploring the learned latent space. Here we use it as a form of style transfer to generate replicate spectra that have been moved in feature space towards the mean spectra of a given class (Figure 21). The production of replicates interpolated towards the mean spectral value for all classes allows us to compare the ensemble classification techniques of the individual class models with that of the all class model. In order to interpolate the latent vector of a sample towards a classes mean value, first the spectral mean is calculated for a class before feeding it through the trained MRA model for a latent vector to be predicted. Interpolation between these two latent vectors occurred with step sizes of $1/10^{th}$. For example at step 0 the interpolated vector matches that of the replicate samples MRA predicted vector, with step 10 seeing it match the classes mean, with each interstitial step moving the latent vectors values from that of the replicate sample towards that of the class means.

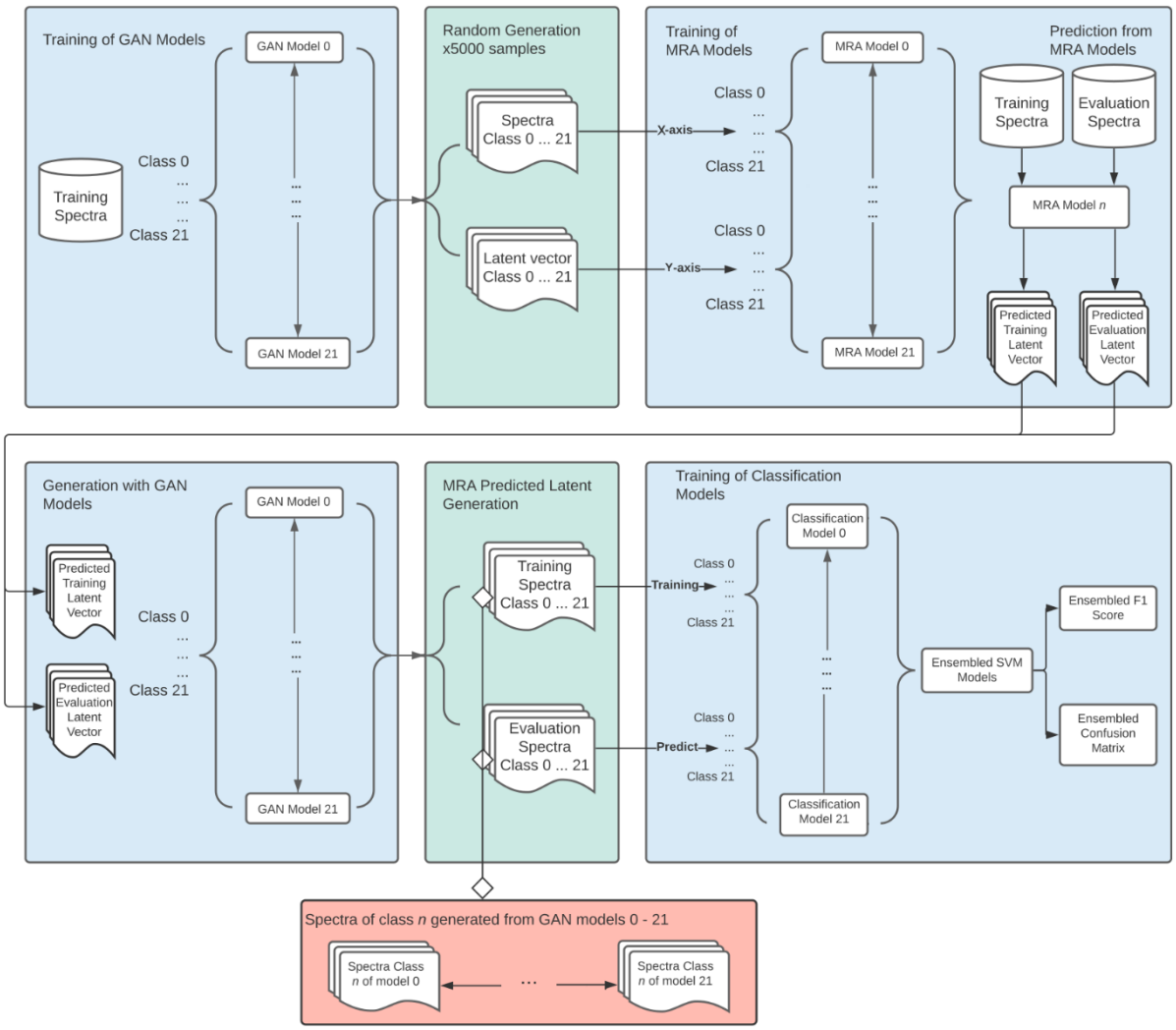


Figure 20. Workflow for stylized replicate generation and ensemble classification.

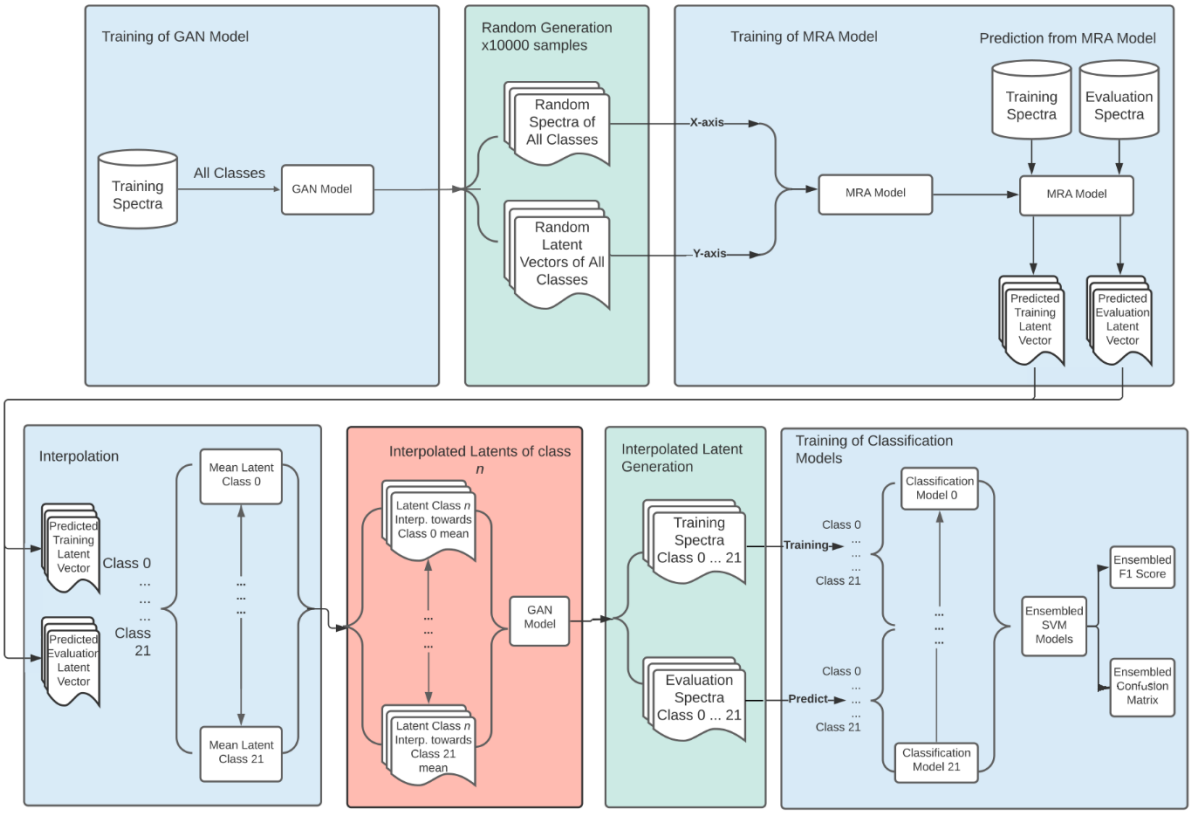


***Figure 21**. Workflow for stylized replicate generation through latent interpolation and ensemble classification.*

All code was written and executed in Python 3.7. The CramérGAN based upon Song (2017) using the Tensorflow 1.8 framework. Support Vector Machine (SVM), and Random Forests (RF) classifiers make use of the Scikit-Learn 0.22.2 library, with Tensorflow 1.8 utilised for the neural network (NN) classifier. Additionally, Scikit-Learn 0.22.2 provided the dimensionality reduction functions for Principal Components Analysis (PCA) and t-distributed Stochastic Neighbourhood Embedding (t-SNE), with Uniform Manifold Approximation and Projection (UMAP) being a standalone library. Hyperparameters for all functions are provided in Appendix B.

4.3. Datasets

The New Zealand (NZ) dataset used in this study is a sub-sample of hyperspectral spectra for 22 species taken from a dataset of 39 native New Zealand plant spectra collected from four different sites around the North Island of New Zealand and made available on the SPECCHIO database (Hueni, Chisholm et al. 2020). These spectra were acquired with an ASD FieldSpecPro spectroradiometer at 1 nm sampling intervals

between 350 – 2500 nm. Following acquisition from the SPECCHIO database, spectra were resampled to 3 nm and noisy bands associated with atmospheric water absorption were removed (1326 – 1464, 1767 – 2004, 2337 – 2500 nm) resulting in 540 bands per spectra. Eighty percent of samples per class were used for training the GAN and 20 % held aside to evaluate classifier performance (Table 11).

Table 11. Plant species classes, training and evaluation sample numbers for New Zealand dataset.

Class ID	Common name	Botanical name	Training samples	Evaluation samples
NZ0	Manuka	*Leptospermum scoparium*	58	14
NZ1	Pohutukawa	*Metrosideros excelsa*	32	8
NZ2	Koromiko	*Hebe stricta*	42	10
NZ3	Lemonwood	*Pittosporum eugenioides*	46	12
NZ4	Kawakawa	*Macropiper excelsum*	34	9
NZ5	Whiteywood	*Melicytus ramiflorus*	48	12
NZ6	Totara	*Podocarpus totara*	34	8
NZ7	New Zealand Flax	*Phormium tenax*	36	9
NZ8	Akiraho	*Olearia paniculata*	8	2
NZ9	Rata	*Metrosideros robusta*	9	2
NZ10	Ngaio	*Myoporum laetum*	38	10
NZ11	Mapou	*Myrsine australis*	36	9
NZ12	Cabbage tree	*Cordyline australis*	32	8
NZ13	Karaka	*Corynocarpus laevigatus*	34	9
NZ14	Kauri	*Agathis australis*	15	3
NZ15	Silver fern	*Cyathea dealbata*	28	7
NZ16	Tangle fern	*Gleichenia dicarpa var. alpina*	14	4
NZ17	Black tree fern	*Cyathea medullaris*	18	4
NZ18	Pigeonwood	*Hedycarya arborea*	18	5
NZ19	Rangiora	*Brachyglottis repanda*	12	3
NZ20	Karamu	*Coprosma robusta*	13	3
NZ21	Red Pine	*Dacrydium cupressinum*	16	4

4.4. Results & Discussion

4.4.1 Regressed latent vector

In order to evaluate the regressed latent vectors ability to reproduce the spectrum of a real sample we examine the average Euclidean and spectral index distances of all training samples for each regression feature step. In general, based on these metrics the regressed latent vectors from an MRA with more independent features generate spectra that better match their real counterpart (Table 12 and Appendix B boxplots). Although the values were often close those generated with a regression feature step of

5 giving a feature length of 108 generally produced the most accurate replication for both all class and single class generator models. Additionally, the reduction from 540 features down to 108 for the independent axis produced a significant increase in speed for training and making predictions from the MRA model. Based on this, all replicated spectra in subsequent sections were produced with a regression feature step size of 5.

Table 12. Step size with lowest mean absolute difference between replicated and real samples for each similarity metric, averaged across all classes.

MODELS	EUC	mSR705	anth_RI	cel_abs	dif_LAI	d_water_stress	nd_leaf_water
KERAS2-ALL-C39	5	5	10	10	5	5	5
KERAS2-C1model	5	5	5	5	1	5	1

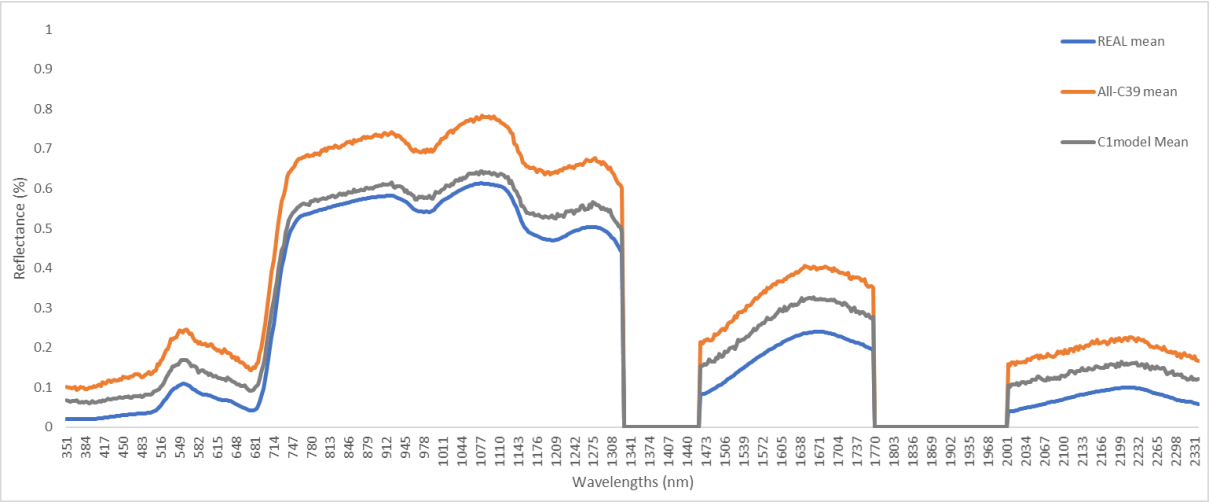


Figure 22. Mean spectra of class 0 for real spectra, all class replicates, and individual class replicates generated with a feature step size of 5.

4.4.2 Training classification ability

In order to examine the classification potential of the replicated spectra SVM, RF and NN classifiers were trained with varying permutations of training and evaluation data subsets (Table 13). The neural network classifier outperformed all others for every combination of data, with SVM being the second most accurate, followed by RF. Classifiers trained then evaluated on the real dataset offer a baseline for comparison with accuracies in accordance with those previously achieved with this dataset (Hennessy, Clarke et al. 2021). All classifiers trained with the real spectra then evaluated on subsets replicated with the all class model or individual class models perform significantly worse, indicating a poor match between the training and

evaluation datasets and therefore a poor replication of the evaluation dataset. The inverse of this is also seen to be true with classifiers trained on replicated spectra failing to classify the real evaluation spectra with acceptable accuracy.

Whilst the replicated spectra appear visually similar to the real spectra (Figure 22) these initial classifications clearly show the replicated samples, either as a training or evaluation subset are mismatched to the real samples distribution. This is further shown by replicated data's ability to classify itself. Replicates generated from the all class GAN model achieves classification accuracies only slightly below that of the real data's baseline, with the NN classifier managing 83.77 %, 1.66 % below the real spectra 85.43 %.

Similarly this is seen when the classifiers are trained and evaluated with replicates from the single class GAN models as well. All classifiers perform extremely well when trained and evaluated on replicated datasets generated from their respective class GAN models demonstrating an increase of 15 – 20 % over the real dataset baseline, with the NN classifier even achieving 100 % accuracy. This of course introduces a significant data leak in that the class labels for the replicated evaluation subset must be used in order to determine which GAN model to generate them from. However, this does show the classification potential of the replicated spectra.

Table 13. Classification accuracies (%) for classifiers trained on real or synthesised spectral data and evaluated on either real or synthesised data for New Zealand datasets based on real class sample sizes.

Train - Evaluation	SVM	RF	NN
Real - Evaluation	82.80	55.39	85.43
Real – Replicate Eval. (All)	17.62	05.98	24.55
Real – Replicate Eval. (single)	18.38	06.16	25.36
All class model			
Replicate Train – Replicate Eval.	75.67	52.09	83.77
Replicate Train – Eval.	13.61	09.36	19.15
Single class model			
Replicate Train – Replicate Eval.*	98.03	75.26	1.0
Replicate Train – Eval.	09.14	04.31	11.02
Replicate Train – Replicate Eval.**	39.94	34.12	40.32

* classifications using replicates generated from their respective class GAN
** classifications using replicates selected based on their similarity to their real sample

In an attempt to eliminate the use of the replicated evaluation subsets class labels every evaluation sample was generated from each of the 22 class GAN models following the reconstruction of their latent vectors. This produced 22 variants of each sample from which one can be selected to represent each evaluation sample. Following testing of numerous distance based metrics, correlation and cosine were found to produce the best results. Clear differences in shape and albedo can been seen between replicates of the same sample generated from GAN model 0 and GAN model 1 (Figure 23). Whilst the replication processes generally achieves accurate replication of spectral shape, albedo is often higher in the replicate than the real sample. To counter this both the replicate and real sample are normalized allowing the distance metrics to focus on shape rather than overall reflectance. Similarly, taking the 1st and 2nd derivates of the spectra highlight the changes in slope and changes in the slopes slope respectively, with the 2nd derivative ultimately providing the greatest distinction between replicates generated from each GAN model (Figure 23). The highest correct assignment to class model was 44.24% based on 2nd derivate samples with both correlation and cosine distance metrics providing the same results. However, using these selections achieved a classification accuracy of only around 40% (table 13**).

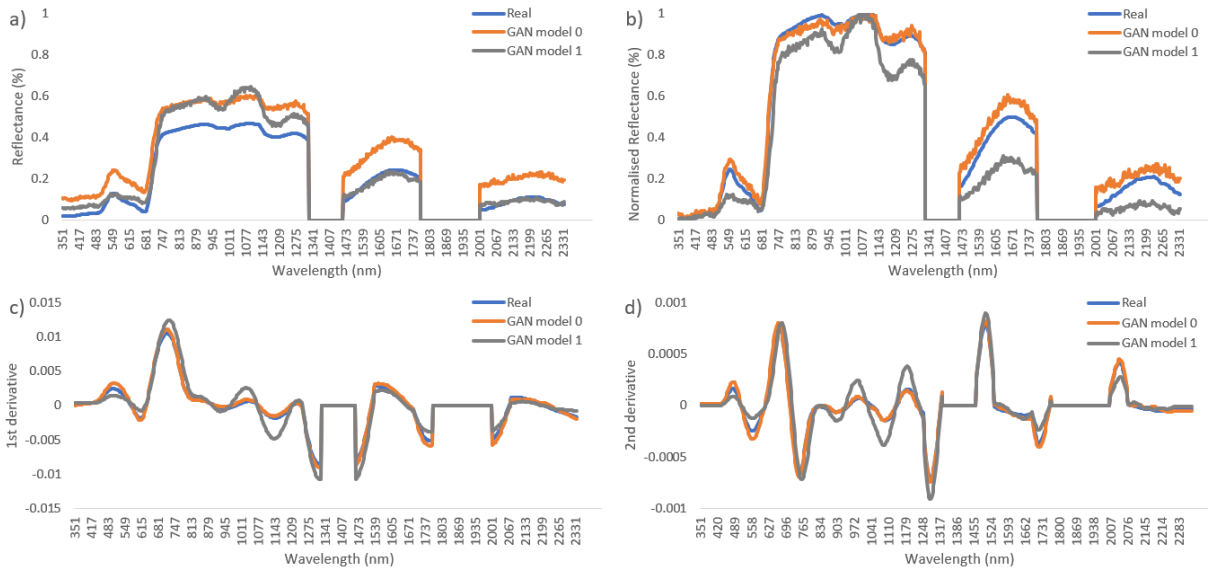


Figure 23. Comparison between real sample from class 0 and replicates of it generated by GAN model 0 and GAN model 1 a) standard reflectance as generated by the GAN, b) albedo normalised, c) 1st order derivative, d) 2nd order derivates.

4.4.3 Ensemble classification

Along with the previously generated variants of each evaluation sample, 22 variants of each sample from the training dataset were also generated in order to train and evaluate the classifiers on fully transformed datasets. All classifiers were trained with a replicate of every sample from the same GAN model variant before being evaluated on the evaluation subset generated from the same GAN model (Figure 20). Whilst never reaching the classification accuracy of the previous section utilizing the class labels of the evaluation subset to select which GAN model to select them from, accuracies were overall very high, independent of which GAN model generated the replicates (Table 14). The lowest accuracies achieved for each classifier all occurred from the same GAN model which had trained only with the real training subset of class 4, 34 samples in total. Whilst being the lowest they still achieved accuracies similar to those when the classifiers were trained and evaluated on the real data subsets. The highest accuracies seen came from GAN model 19 for the RF and NN classifiers, though from GAN model 12 for SVM, with model 19 being its second highest accuracy. These high classification accuracies represent 10-20 % increases over those produced by the real – real classifiers.

An even greater increase in accuracy was seen after ensembling the predictions from each of the 22 variant classification models for each of the classifiers. Interestingly, although each of the individual results differed between the SVM and NN classifiers their final ensembled accuracy was the same at 98.62 %, a dramatic increase of 15.82 % and 13.19 % respectively over the original real spectra classifications. While only reaching a comparative low of 85.81 % the ensembled RF classifier saw an improvement of 30.42 %.

Table 14. Classification accuracies (%) for each individual classification and ensembled classification for replicates generated from individual class GAN models.

Train - Evaluation	SVM	RF	NN
GAN model 0	87.49	53.81	87.18
GAN model 1	89.84	44.57	82.50
GAN model 2	90.14	50.70	92.28
GAN model 3	92.63	50.25	91.14
GAN model 4	**85.98***	**55.65***	**82.32***
GAN model 5	89.26	56.99	90.34
GAN model 6	92.78	56.47	90.00
GAN model 7	93.71	67.85	95.46
GAN model 8	93.44	59.64	94.09
GAN model 9	93.40	70.98	92.18
GAN model 10	90.52	70.14	93.36
GAN model 11	94.18	54.67	93.98
GAN model 12	**97.33****	59.02	93.96
GAN model 13	94.70	62.59	94.05
GAN model 14	95.77	69.15	94.49
GAN model 15	94.31	62.30	94.17
GAN model 16	94.50	64.49	94.40
GAN model 17	95.33	73.52	93.97
GAN model 18	89.23	57.13	95.48
GAN model 19	96.69	**75.03****	**97.38****
GAN model 20	94.74	63.22	93.17
GAN model 21	93.49	70.87	94.55
Ensembled	98.62	85.81	98.62

* Lowest achieved accuracy for each classifier
** Highest achieved accuracy for each classifier

Such a style transfer-like post processing method is only made possible because there are individual GAN models for each class. In order to test a similar method and ensemble classification on the single GAN model trained on all classes an interpolation of latent vectors was used. The optimum level of interpolation was found to be 3 steps towards the mean of each class, resulting in only a small style transformation of the

replicated spectra. Interpolated latents were produced for each training and evaluation sample before being generated by the all class model (Figure 21). As with the previous ensemble method each classifier was trained and evaluated with variants of each class before their individual predictions were ensembled together (Table 15). Increases of ~5 % were seen for the SVM and RF ensembled classifications over the all class model classifications on the non-style transferred replicates raising from 75.67 To 80.49 % and 55.09 To 60.32 % respectively. However, the NN classifier performed poorly with individual variant results ranging from 48.85 % to 61.98 % resulting in an ensembled accuracy of 60.97 % only slightly higher than that of the RF. Whilst having the benefit of only requiring the training and generation from a single GAN model and producing higher accuracies than the previous best for the replicated spectra from the single all class model this method ultimately did not outperform the original spectra.

Table 15. Classification accuracies (%) for each individual classification and ensembled classification for replicates generated from latent interpolation of the all classes GAN model.

Train - Evaluation	SVM	RF	NN
Class 0	77.23	55.15	50.45
Class 1	**68.68***	**44.18***	51.79
Class 2	77.00	55.04	48.93
Class 3	76.93	53.41	52.83
Class 4	76.20	55.80	56.81
Class 5	78.08	54.68	**48.85***
Class 6	79.70	54.11	55.33
Class 7	76.72	55.12	56.40
Class 8	72.53	52.77	52.82
Class 9	72.96	53.33	56.46
Class 10	78.67	57.70	53.76
Class 11	79.72	56.73	55.26
Class 12	80.39	**60.73****	59.26
Class 13	78.23	58.83	56.27
Class 14	79.93	59.91	57.96
Class 15	**80.99****	55.62	58.23
Class 16	79.08	59.96	58.08
Class 17	77.44	54.73	54.98
Class 18	78.88	55.14	**61.98****
Class 19	74.78	55.81	56.61
Class 20	79.49	60.22	58.81
Class 21	80.15	56.94	59.72
Ensembled	80.49	60.32	60.97

* Lowest achieved accuracy for each classifier
** Highest achieved accuracy for each classifier

4.5. Increased classification potential of replicate spectra

Whilst the ensembled classifications returned the highest accuracy, the individual classifications that create it performed as well as, or better than, classifications of the original real spectra. This would indicate the process of replicating the training and evaluation subsets inherently improves classification power. Focusing on replicates from GAN models 4 and 19 which produced the lowest and highest accuracies respectively we examine the differences between replicated and real spectra and attempt to interpret the mechanism behind this.

Table 16. Confusion matrix of classification accuracy from SVM classifier trained and evaluated on replicates generated from GAN model 19.

CLASS	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	TOTAL	PRODUCERS ACCURACY
0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**14**	100
1	1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	87.5
2	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	100
3	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**12**	100
4	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**9**	100
5	0	1	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**12**	91.7
6	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	100
7	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**9**	100
8	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	100
9	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	**2**	100
10	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	**10**	100
11	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	**9**	100
12	0	0	1	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	**8**	87.5
13	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	**9**	100
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	**3**	100
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	**7**	100
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	**4**	100
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	**4**	100
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	**5**	100
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	**3**	66.7
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	**3**	66.7
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	**4**	100
TOTAL	**15**	**8**	**11**	**12**	**9**	**11**	**8**	**9**	**2**	**2**	**10**	**9**	**7**	**10**	**3**	**7**	**4**	**5**	**5**	**2**	**2**	**4**	**155**	
USERS ACCURACY	93.3	87.5	90.9	100	100	100	100	100	100	100	100	100	100	90	100	100	100	80	100	100	100	100		

Table 17. Confusion matrix of classification accuracy from SVM classifier trained and evaluated on replicates generated from GAN model 4.

CLASS	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	TOTAL	PRODUCERS ACCURACY
0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**14**	100
1	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	**8**	87.5
2	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**10**	100
3	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**12**	100
4	2	1	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**9**	66.7
5	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**12**	100
6	1	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**8**	87.5
7	0	0	0	0	0	0	0	8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	**9**	88.9
8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	50
9	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**2**	0
10	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	**10**	100
11	0	0	0	1	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	**9**	88.9
12	0	0	0	0	1	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	**8**	87.5
13	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	**9**	100
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	**3**	100
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	1	0	0	**7**	85.7
16	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	**4**	50
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	**4**	100
18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	**5**	80
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	**3**	33.3
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	**3**	33.3
21	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	**4**	75
TOTAL	**19**	**9**	**12**	**14**	**7**	**12**	**8**	**8**	**1**	**0**	**10**	**8**	**7**	**10**	**3**	**9**	**2**	**4**	**5**	**2**	**1**	**4**	**155**	
USERS ACCURACY	73.7	77.8	83.3	85.7	85.7	100	87.5	100	100	0	100	100	100	90	100	66.7	100	100	80	50	100	75		

The differences in classification accuracy are highlighted by the users and producers accuracies in the confusion matrix for the replicates generated from model 4 and 19 (Table 16, 17). Replicates that were erroneously classified differed between the two subsets without being concentrated to any particular classes. To further investigate the differences between the replicates of these two subsets we focus upon class 4 which can be seen to produce 3 errors of omission and 1 of commission when generated by model 4 and to produce no errors when generated by model 19.

The mean spectra for both the training and evaluation subsets form two distinct groupings based on the GAN model they were generated from demonstrating the significant style transfer effect. The mean spectra replicated for classes 0, 1, 4 and 6 from model 4 more closely match the real mean spectra of class 4 than those replicated from model 19 do for the real mean spectra of class 19 (Figure 24). Despite the clear differences between the generation models there do not appear to be differences between the mean of the training and evaluation subsets that would explain the improved classification performance of those generated from model 19 over those from model 4 or the real spectra.

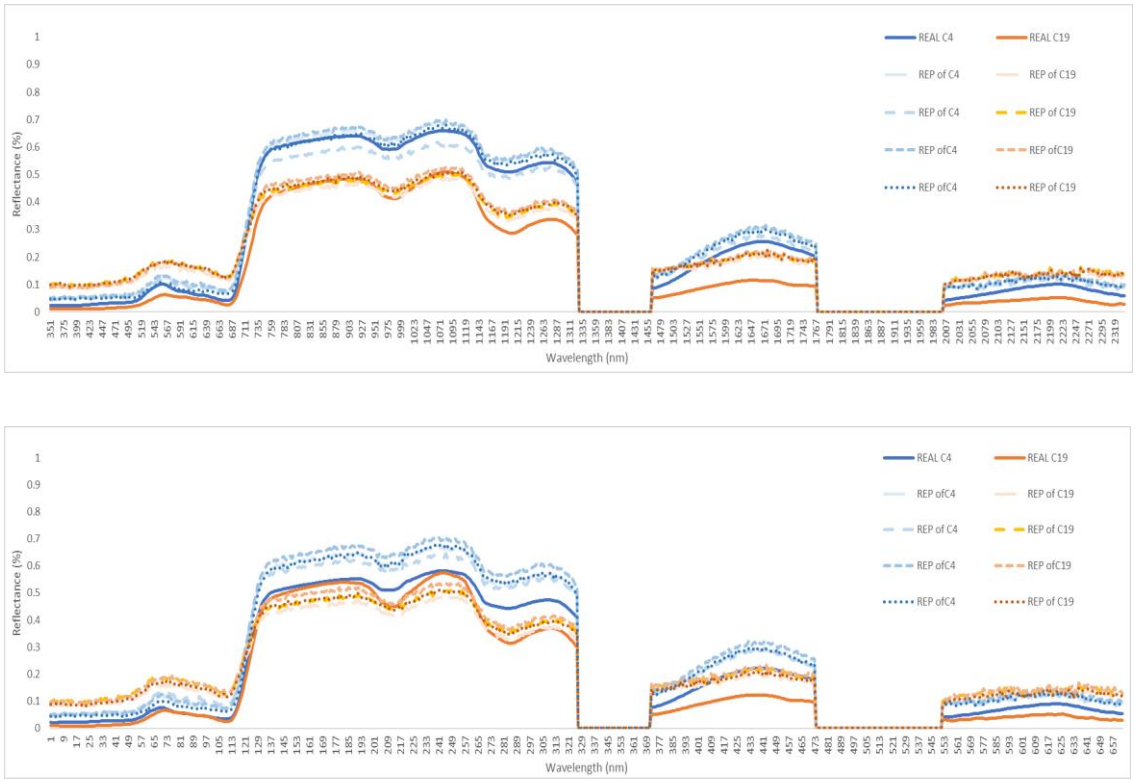


Figure 24. Top) training spectra, Bottom) evaluation spectra. Real mean of c4, real mean of c19, replicate mean from its own model, replicate mean ofc4 and replicate mean of c19 for classes 0,1,4, and 6.

Dimensionality reduction however does provide insight into the differing classification potential of each replicate subset. Clear differences in intra and inter-class relationships in both the training and evaluation subsets are identified following the plotting of the UMAP reduced spectra (Figure 25). Replicates generated from model 19 demonstrate clear tight clusters for each class with minimal outliers. While the clustering of model 4 replicates visually appears significantly less obvious, though classes do appear more organised compared to the real samples. However, this is a potential visual artifact of the overall reduced variance of the subset with x/y spreads of 10/14 units for model 4 replicates and 45/50 for the real spectra. This is further backed up with model 4 replicates and real spectra producing very similar classification accuracies with 85.98 % and 85.43 % being the highest recorded respectively, indicating similar class relationships.

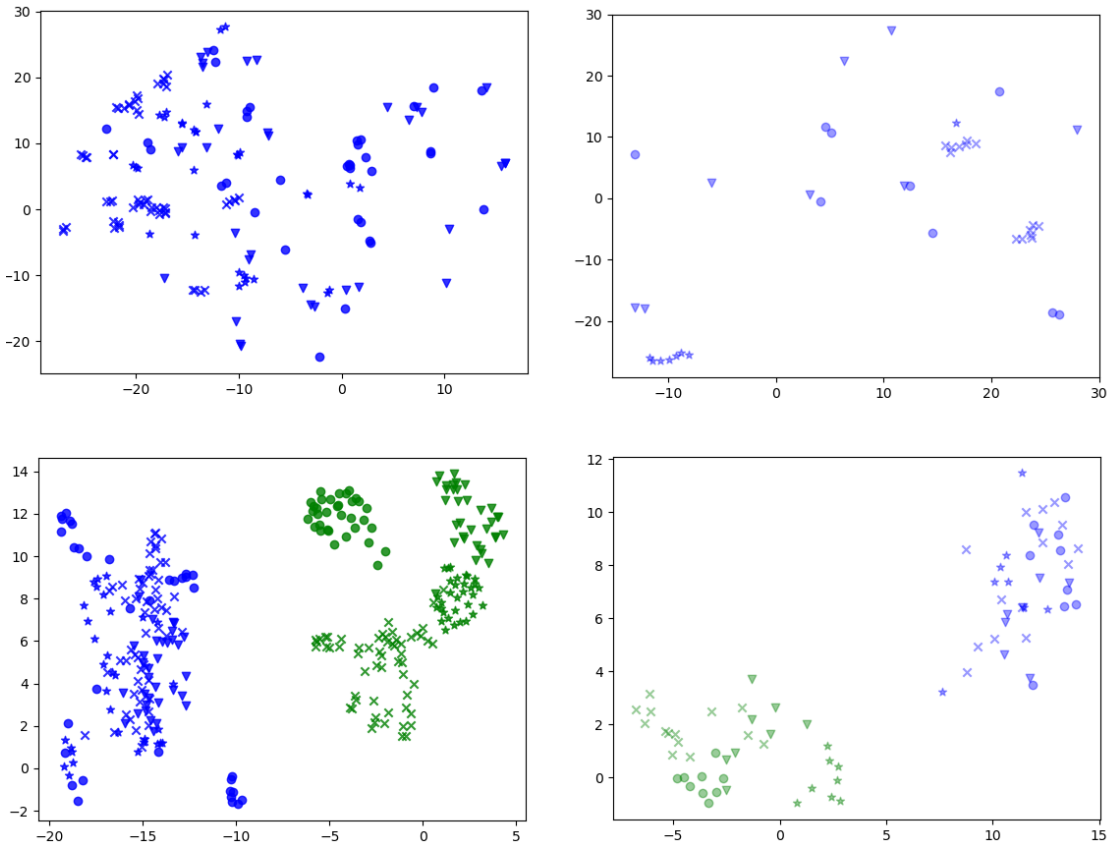


Figure 25. UMAP transformed representation x = class0, * = class1, o = class4, v = class6. Top left) real spectra training subset, top right) real spectra evaluation subset, bottom left) replicate spectra training subsets, bottom right) replicate spectra evaluation subsets (Blue model 4)(Green model 19).

4.6. Conclusions

In this paper we have demonstrated the ability to replicate hyperspectral samples of numerous New Zealand plants by use of a generative adversarial network. Training classifiers with these replicated spectra produced large increases in classification accuracy. We believe the driving force for this increase to be a type of style transfer that occurs when generating a spectrum from a model trained on a different class. This results in the replicated spectrum being imprinted with attributes of the GAN models training class, ultimately creating a hybrid class with differing intra and inter-class relationships compared to its real spectra. We have seen these hybrid classes display similar classification potential as the real spectra, achieving accuracies of 85.98 % compared to 85.43 % for real spectra. Though some of these hybrid classes such as those from GAN model 19 show reduced intra-class variance and clear inter-class differentiation producing a maximum accuracy of 97.38 % with a neural network classifier. However, it is not necessary to select for the best performing GAN model

replicates, rather simply classifying all and ensembling the results. Not only does this remove the requirement of selecting a single generative model, in this case it resulted in an increase of classification accuracy over any individual classification. Both the support vector machine and neural network classifiers produced an ensembled accuracy of 98.62 %, an improvement of 13.19 % over the real spectral samples. Future work with this technique is required to demonstrate generalizability to other datasets, both hyperspectral and non-hyperspectral as well as vegetation and non-vegetation. However, this work represents an initial large step towards a methodology that can overcome intra and inter-class issues associated with supervised classification of vegetation HS data.

4.7. References

Alharbi, Y. and P. Wonka (2020). “Disentangled image generation through structured noise injection.” In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Castro-Esau, K. L., G. A. Sánchez-Azofeifa, B. Rivard, S. J. Wright and M. Quesada (2006). "Variability in leaf optical properties of Mesoamerican trees and the potential for species classification." *American Journal of Botany* 93(4): 517-530.

Galvao, L. S., A. R. Formaggio and D. A. Tisot (2005). "Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data." *Remote Sensing of Environment* 94(4): 523-534.

Gatys, L. A., A. S. Ecker and M. Bethge (2016). “Image style transfer using convolutional neural networks.” In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition.

Gitelson, A. A., M. Merzlyak, Y. Zur, R. Stark and U. Gritz (2001). "Non-destructive and remote sensing techniques for estimation of vegetation status." In: Proceedings of the 3rd European Conference on Precision Agriculture (ECPA): 205–210

Hennessy, A., K. Clarke and M. Lewis (2021). "Generative adversarial network synthesis of hyperspectral vegetation data." *Remote Sensing* 13(12): 2243.

Hueni, A., L. A. Chisholm, C. Ong, T. J. Malthus, M. Wyatt, S. A. Trim, M. E. Schaepman and M. Thankappan (2020). "The SPECCHIO Spectral Information System." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 5789-5799.

Jing, Y., Y. Yang, Z. Feng, J. Ye, Y. Yu and M. Song (2019). "Neural style transfer: A review." *IEEE Transactions on Visualization and Computer Graphics* 26(11): 3365-3385.

Le Maire, G., C. François, K. Soudani, D. Berveiller, J.-Y. Pontailler, N. Bréda, H. Genet, H. Davi and E. Dufrêne (2008). "Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass." *Remote Sensing of Environment* 112(10): 3846-3864.

Lopinto, E. and C. Ananasso (2020). “The Prisma hyperspectral mission.” In: Proceedings of the 33rd EARSeL Symposium, Towards Horizon.

Nagler, P., C. Daughtry and S. Goward (2000). "Plant litter and soil reflectance." *Remote Sensing of Environment* 71(2): 207-215.

Nicolas Audebert, B. L. S., Sébastien Lefèvre (2018). "GAN for realistic synthesis of hyperspectral samples." IGARSS 2018 IEEE International Geoscience and Remote Sensing Symposium.

Pu, R., S. Bell, L. Baggett, C. Meyer and Y. Zhao (2012). "Discrimination of seagrass species and cover classes with in situ hyperspectral data." *Journal of Coastal Research* 28(6): 1330-1344.

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

Song, J. (2017). cramer-gan. https://github.com/jiamings/cramer-gan, GitHub.

Woodcock, C. E. (2002). "Uncertainty in remote sensing." *Uncertainty in Remote Sensing and GIS*: John Wiley & Sons, 19-24.

Intentionally left blank

CHAPTER FIVE: Discussion and Conclusions

5.1. Overview

This thesis has focused upon intra and inter-class variability of vegetation hyperspectral reflectance spectra and its detrimental effects on hyperspectral classification accuracy and generalisability. Increased intra-class variance can decrease the chance that labelled training spectra are representative of the entire class and often requires the collection of more labelled spectra to ensure the full spectral distribution of classes is represented. At the same time, increased intra-class variance also decreases inter-class distance, dramatically increasing the chance of class overlap and therefore classification error. The overarching aim of this thesis was to develop methods of improving the ratio between intra and inter-class variability in hyperspectral vegetation spectra, and ultimately increasing classification accuracy, reliability and generalisability. This was addressed in three ways: (1) reviewing the hyperspectral classification literature of the past two decades, while also performing pre-processing and classification trials on a hyperspectral dataset, (2) increasing the number and distribution of classifier training samples through data augmentation, and (3) improving the intra/inter-class relationship with deep generative sample transformation.

5.2. Key Outcomes

Chapter two examined feature/waveband selection in the hyperspectral vegetation classification literature. Such a comprehensive review of spectral feature selection methods and resultant waveband selections had not yet been performed. Additionally, the findings of the literature review were supported by feature selection trials with varying parameters to further investigate the impact of study design on waveband selection and classification accuracy. The use of feature selection was found to have three main drivers; the need to reduce data size and ease computational requirements; to aid in class differentiation by selecting wavebands that minimised intra-class variance whilst maximising inter-class variance; though it is primarily used to reduce the number of discriminatory features relative to sample size to meet the requirements of most statistical classifiers. Having more discriminatory variables than sample size violates many of the statistical requirements of such classifiers, hence reducing the number of features used could improve classification accuracy. This chapter identified that there were many parameters of plant spectral discrimination studies that

influenced which spectral features were selected, the primary of which being simply the feature selection method that was implemented. Whilst all other differences between studies such as taxonomic, structural and functional groups of the target samples, and the methods and scale at which hyperspectral measurements were recorded all influenced feature selection, their influence on which wavebands were selected for discrimination and classification was overshadowed by the feature selection method. Despite the strong influence of the feature selection methods, some important and consistent patterns of feature selection were observed throughout the literature. Notwithstanding their ubiquitous and conserved nature visible wavelengths associated with photosynthetic pigments have been known to play an important role in discriminating vegetation classes: this was clear by the increased selection rate of the visible region. Similarly, red edge wavelengths were consistently selected. In contrast, the wavebands from the NIR have been reported as significant for classification, though the findings of this chapter challenge that. While NIR wavebands were selected their selection was almost entirely driven by the use of the SDA feature selection method. However, the use of SDA as a feature selection method has been heavily criticised in the statistical and data science literature. Perhaps the clearest outcome of this chapter was that no single spectral feature selection method outperformed the others, with significant variation in waveband selection between and within studies. Additionally, multiple feature selection subsets were seen in the literature and shown in my own trials to work equally as well on the same data, or in fact, no feature selection at all.

Consequently, an alternative method for handling small sample sizes and attempting to improve intra and inter-class variances was investigated. Chapter three explored data augmentation as such an alternative method. The augmentation of the original hyperspectral vegetation dataset for training classifiers was performed by the generation of new synthesised spectra that visually resembled the spectra from any given class, while respecting the sampled class distribution. This synthesis was achieved by the use of a generative adversarial network, which at the time of designing and performing the experiments of the chapter was a unique and novel method to the hyperspectral literature. Whilst similar work has since been published the generative model developed for this chapter was able to overcome two key issues encountered by

other researchers. The use of the Cramer Distance as the evaluation metric meant that the spectra synthesised in this chapter more closely resembled the real spectra. This was primarily due to not encountering the issue of mode collapse where only a subset of a class's distribution is learned. The addition of synthesised spectra to original spectra, creating an augmented dataset, increased classification accuracy across nearly all trials for two independent hyperspectral plant datasets and three different classification methods. It was also shown that synthetic spectra that came from sparse or outlying regions of class feature spaces contributed more discriminatory power to the classifiers than others. This can be seen as the augmented dataset acting in the same manner as the actual physical recording of new spectral data so that the labelled classifier training subset better captures intra-class variability and therefore potentially better represents spectra within the evaluation dataset. However, this is a double-edged sword with the addition of spectra in sparse and outlying regions of class feature spaces potentially decreasing inter-class distances and therefore increasing the likelihood of class overlap.

Chapter four further explored the potential of these generative models and synthesised spectra attempting to both reduce intra-class variance as well as improve class separation. The previous chapter generated new random spectra from a plausible distribution of a class, whereas the approach taken here was to generate a synthetic replicate of each sample for training and evaluation subsets. Whilst these replicates were not perfect matches for their real counterparts the replicate training spectra achieved a higher classification accuracy on replicated evaluation spectra than the real spectra achieved when classifying the real evaluation data. However, this was only possible with perfect class knowledge of the evaluation dataset so they could be synthesised from the generative model that had been trained on their class. This is something not possible when dealing with unknown samples for classification in a real-world setting. This issue was overcome by synthesising a version of all samples from each generative model trained independently on each class, in this case resulting in 22 versions for every real sample. It was shown that as long as the training and evaluation samples were generated from the same model they performed as well as, or significantly better than, the real samples. Rather than having an additional test subset

to help identify which generative model produces the highest classification accuracy it was found that an ensemble classification model that incorporated 22 individual classifiers, one for each class, produced the highest overall accuracy of all, whilst removing the use of the evaluation subsets labels.

5.3 Significance and Implications of the Research

The methods presented here have the potential to revolutionise the field of hyperspectral classification. They demonstrate a potential to not only increase classification accuracy but, most significantly, they present a possibility to improve the generalisability and transferability of classification models. This provides researchers with the opportunity to mature their research into widescale knowledge and products for environmental mapping and monitoring.

Chapter two highlights the differences in wavebands selected within and between studies and raises questions about the appropriateness of replicating the use of feature selection methods or selected wavebands directly from the literature. The transferability of feature selection is questionable outside of the few broad trends identified, especially so in the case of SDA, which has been used extensively throughout the literature though is perhaps not suitable for the task of spectral feature selection. Additionally, the need for feature selection at all is now uncertain, with some studies showing no real benefit or even deficits in overall accuracy when performing feature selection. Additionally, the bottleneck of computing resources is generally no longer a factor. Furthermore, this thesis has shown that data augmentation can solve the problem of low sample numbers.

Chapter three demonstrated the ability for a generative deep learning algorithm based on adversarial neural networks to successfully learn and synthesise realistic vegetation spectra. The use of these synthesised spectra to augment real spectral datasets allows for the training of classifiers that benefit from large sample numbers without a researcher needing to collect additional labelled spectra from the field. This is of increasing significance as modern machine and deep learning algorithms tend to require larger datasets.

The increases in classification accuracy of chapter three were expanded upon in chapter four, resulting in the most significant outcome of this thesis. The transformed replicates synthesised by the generative models demonstrated dramatic changes to intra-class variance, in best case scenarios acting to tightly cluster all members of a class. This clustering also improved inter-class relationships, creating clear delineations between classes and allowing for all classifiers tested to achieve remarkably increased accuracies, following ensembling of results. Notably the processing pipeline to achieve this does not require any additional spectral samples or metadata, making the method applicable to all historically gathered hyperspectral data, providing increased utility for spectral libraries and shared databases.

5.4 Recommendations for Future Research

In order to fully capitalise on the research presented here the following research should be performed.

The methods presented in chapter four show considerable potential for improving hyperspectral classification with the logical next step being to apply these methods to other hyperspectral vegetation datasets. In particular, further evaluation of the generalisability of my methods is needed when applied to spectral data with high degrees of both spatial and temporal intra-class variability. Potential for mitigation of artefacts and uncertainties added to spectra from sources such as atmospheric correction algorithms should also be explored. Following on from this the method could be extended to non-vegetation hyperspectral data, or potentially non-hyperspectral remote sensing data. Spectra produced by the current implementation of the GAN are visually identifiable as synthetic predominantly by the spectra-wide noise, further refinement of the GAN could potentially improve this, generating spectra indistinguishable from real. Additionally, this noise should be investigated as a potential cause of new correlations increasing uncertainties within the data or potentially act as class specific signals that could influence classification results. As well, the field of deep machine learning is improving at substantial rates, providing new algorithms and techniques that could be used to improve the GAN and resultant synthesised spectra or even present non-GAN based generation methods. The estimation of latent vectors to replicate spectra also has potential for refinement so that replicated spectra generated

from their latents more closely resemble the real spectra. The ability to synthesise hyperspectral data based on a prior sample raises the potential of using these methods to transform multispectral samples into hyperspectral and should be explored. Whilst these methods were designed specifically for hyperspectral vegetation reflectance measurements they are in fact data-agnostic and could be applied to a wide variety of fields. Hence the broader generalisability of these methods to improve classification in other fields should be investigated.

5.5 Conclusions

New, accurate and generalisable methods are required to transform the ever-increasing amount of raw hyperspectral data into actionable knowledge for applications such as environmental monitoring and precision agriculture. The research presented in this thesis provides early steps towards utilising modern deep learning techniques in novel ways to address the simple fact that not all members of the same class closely match each other, and ultimately improving overall classification accuracy of hyperspectral vegetation data. Generative deep learning models for data augmentation is a fledgling field in general and especially so in the discipline of hyperspectral remote sensing. This thesis is one of only a few studies to explore these methodologies for this purpose and I believe to be the only one to use style transfer concepts and to produce such significant increases in classification accuracy. The results presented here appear to significantly reduce the influence of intra and inter-class variability and may offer the key to fast, accurate and generalisable classifications that can handle the volume of hyperspectral data that has and will be produced.

Appendix A

Table a: Architecture and hyperparameters of the neural network classifier trained for 1000 epochs with Kullback-Leibler divergence loss, Adam optimiser with a learning rate of 0.00001, and a batch size of 32

Layer type / parameters	Shape	Activation
Conv1D	(100,100)	ReLU
Conv1D	(100,50)	ReLU
Conv1D	(100,10)	ReLU
Flatten		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	4096	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	22	Softmax

Table b**.** Layer architecture of the GANs discriminator.

Layer type/parameters	Shape	Activation
Dense	1024	Leaky ReLU (alpha = 0.2)
Dense	1024	Leaky ReLU (alpha = 0.2)
Dense	256	Linear

Table c. Epochs with Kullback–Leibler divergence loss, Adam optimiser with a learning rate of 0.00001, and a batch size of 32.

Layer type/parameters	Shape	Activation
Conv1D	(100,100)	ReLU
Conv1D	(100,50)	ReLU
Conv1D	(100,10)	ReLU
Flatten		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	4096	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	22	Softmax

Table d: Hyperparameters used during t-SNE dimension reduction for each dataset.

Dataset	Perplexity	Early exaggeration	Learning rate
INDI	50	5	2
NZ	50	1	15

Table e: Hyperparameters used during UMAP dimension reduction for each dataset.

Dataset	Number of neighbours	Minimum distance	Distance metric
INDI	20	1	Canberra
NZ	100	0.3	Correlation

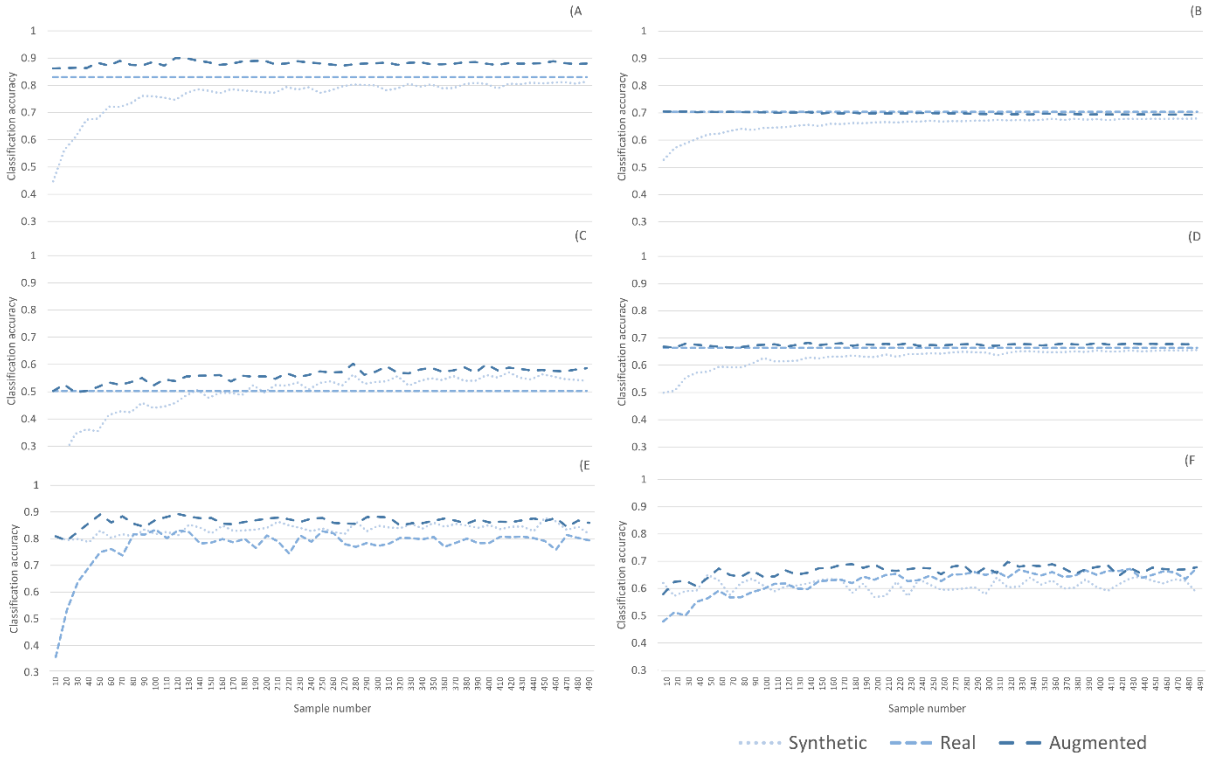
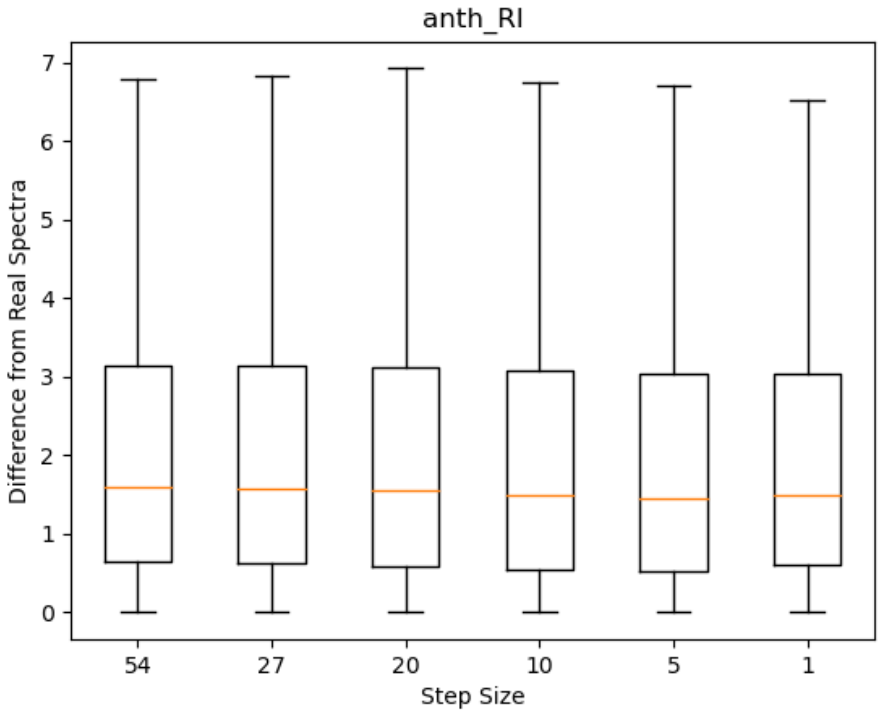


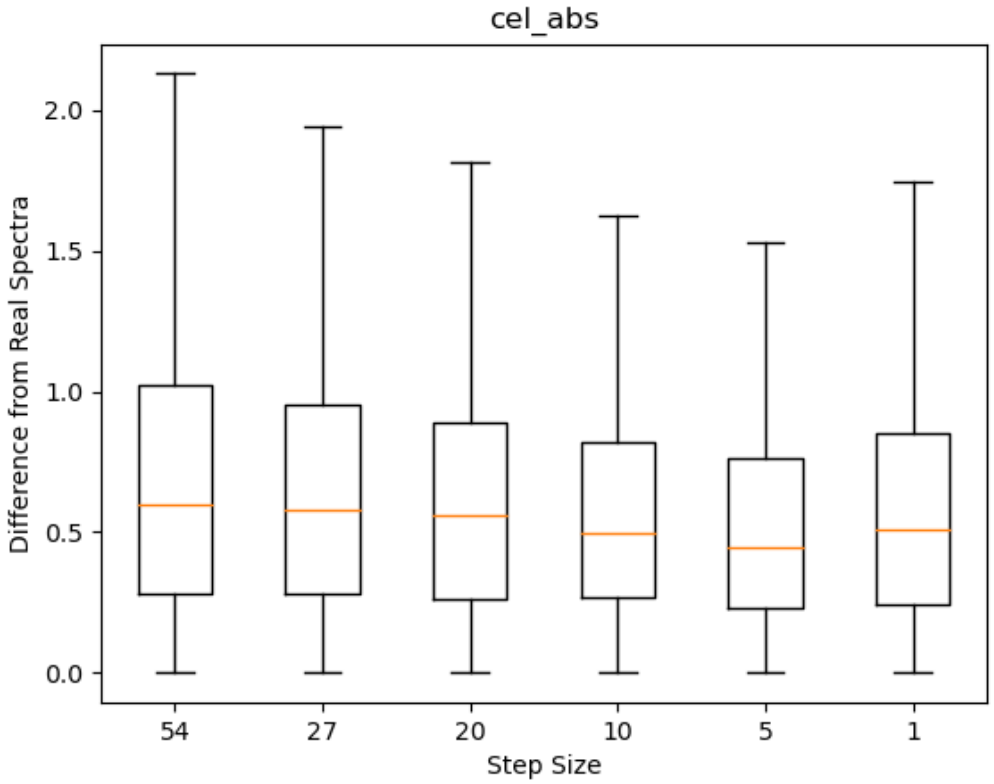
Figure a. Classification accuracies for classifiers trained on real, synthesised, or augmented spectral data and evaluated on evaluation dataset for both Indian Pines and New Zealand datasets ranging from 10 - 490 samples per class. (A New Zealand dataset; SVM classifier, (B Indian Pines dataset; SVM classifier, (C New Zealand dataset; RF classifier, (D Indian Pines dataset; RF classifier, (E New Zealand dataset; NN classifier, (F Indian Pines dataset; NN classifier.

Appendix B

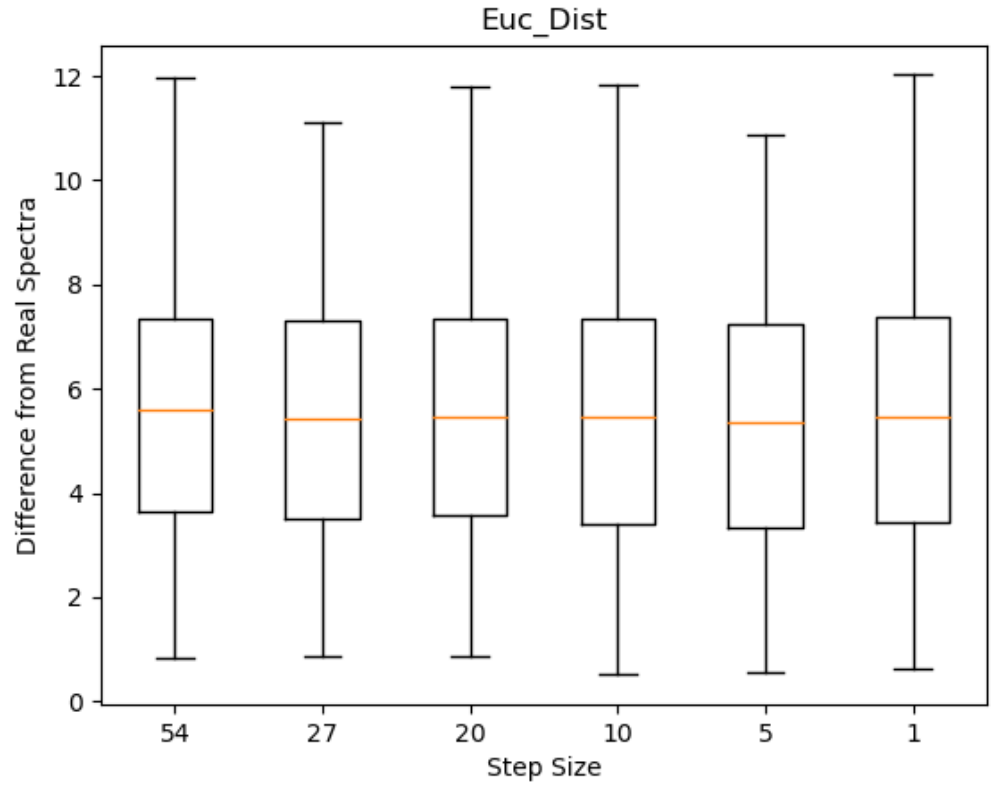
Table a: Architecture and hyperparameters of the neural network classifier trained for 1000 epochs with Kullback-Leibler divergence loss, Adam optimiser with a learning rate of 0.00001, and a batch size of 32

Layer type / parameters	Shape	Activation
Conv1D	(100,100)	ReLU
Conv1D	(100,50)	ReLU
Conv1D	(100,10)	ReLU
Flatten		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	4096	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	2048	Leaky ReLU (alpha = 0.2)
Batch Normalization (momentum = 0.4)		
Dropout (0.5)		
Dense	22	Softmax

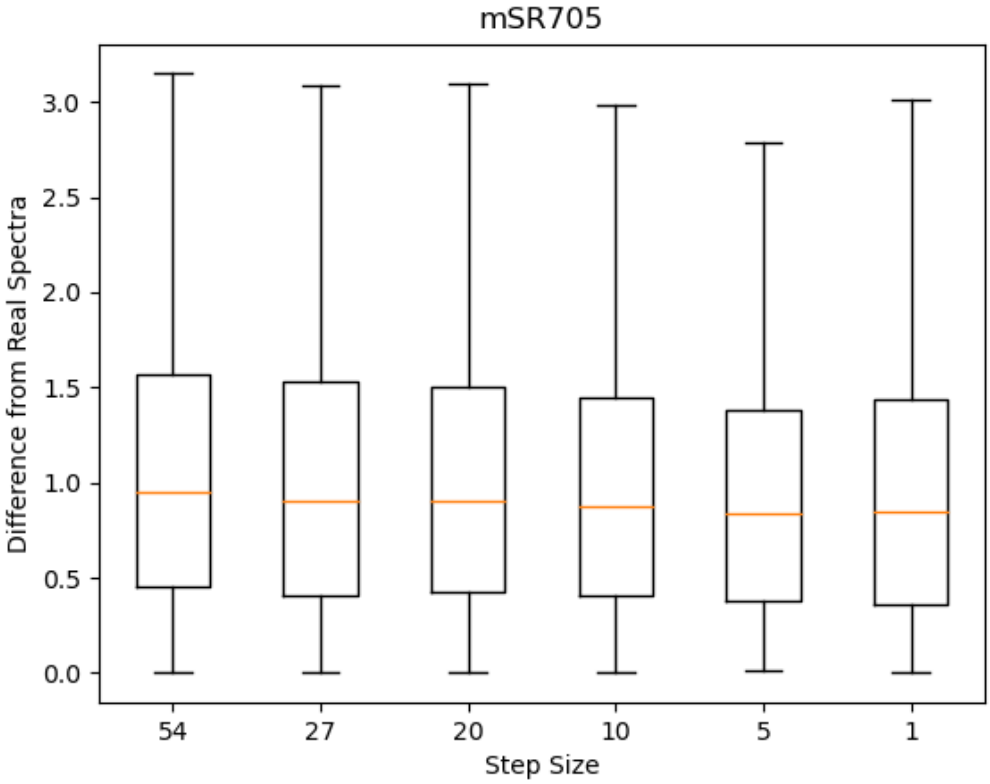



cel_abs
Difference from Real Spectra
2.0
1.5
1.0
0.5
0.0
54
27
20
10
5
1
Step Size

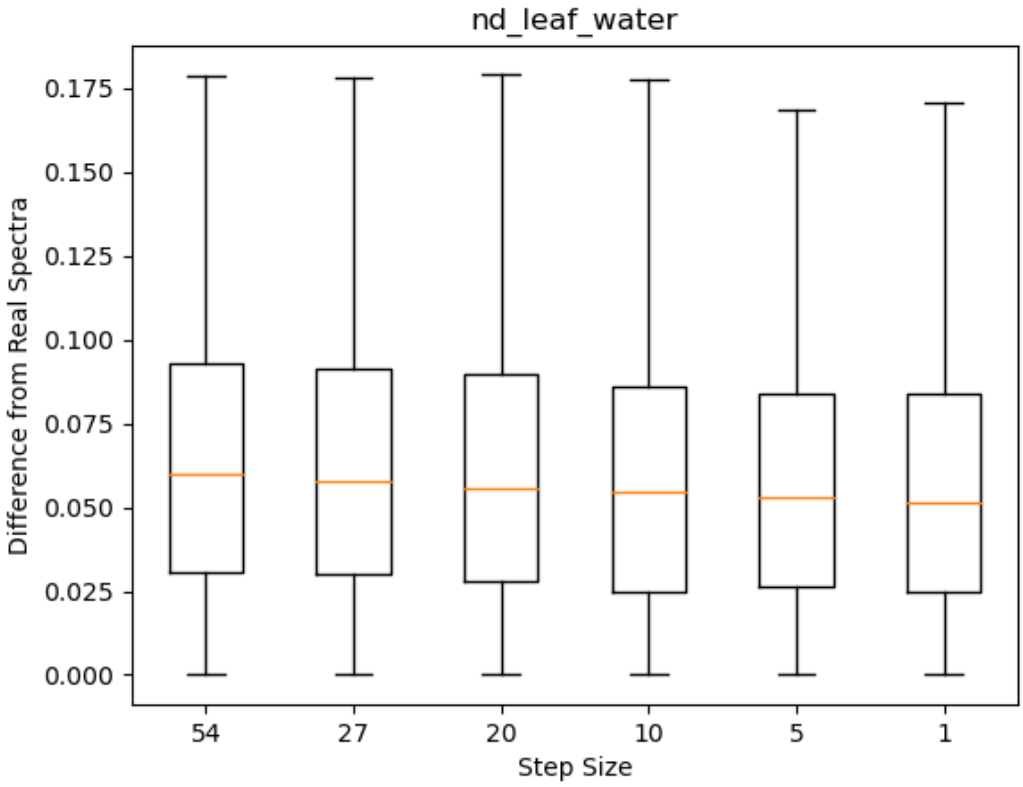
1


Euc_Dist
Difference from Real Spectra
12
10
8
6
4
2
0
54
27
20
10
5
1
Step Size

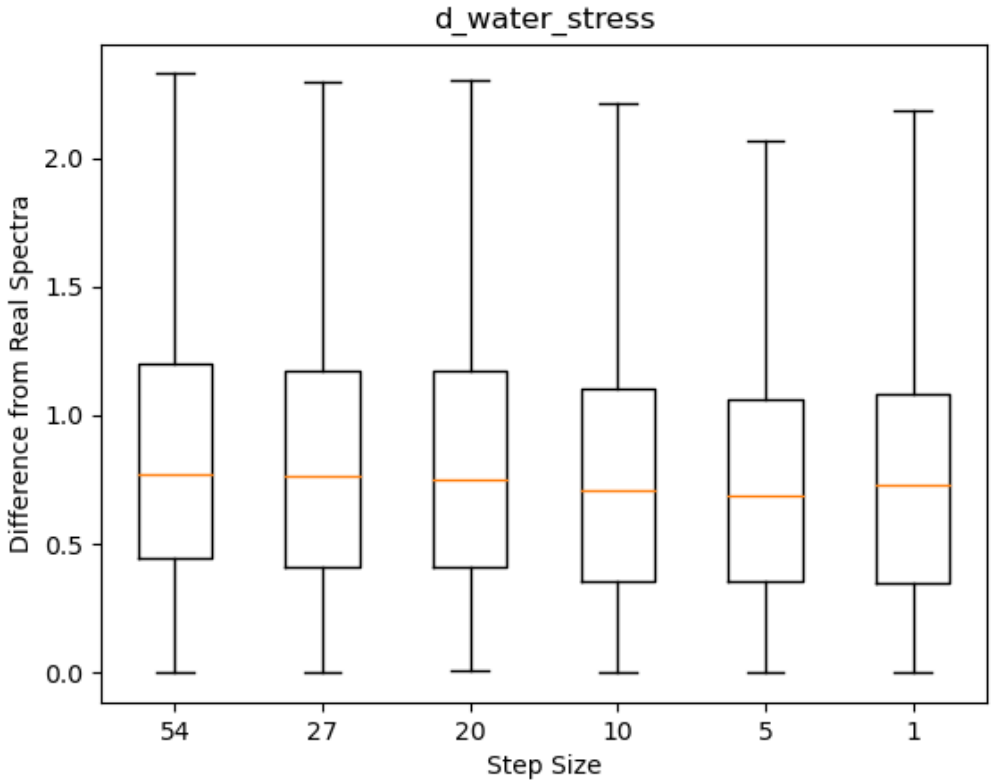
2

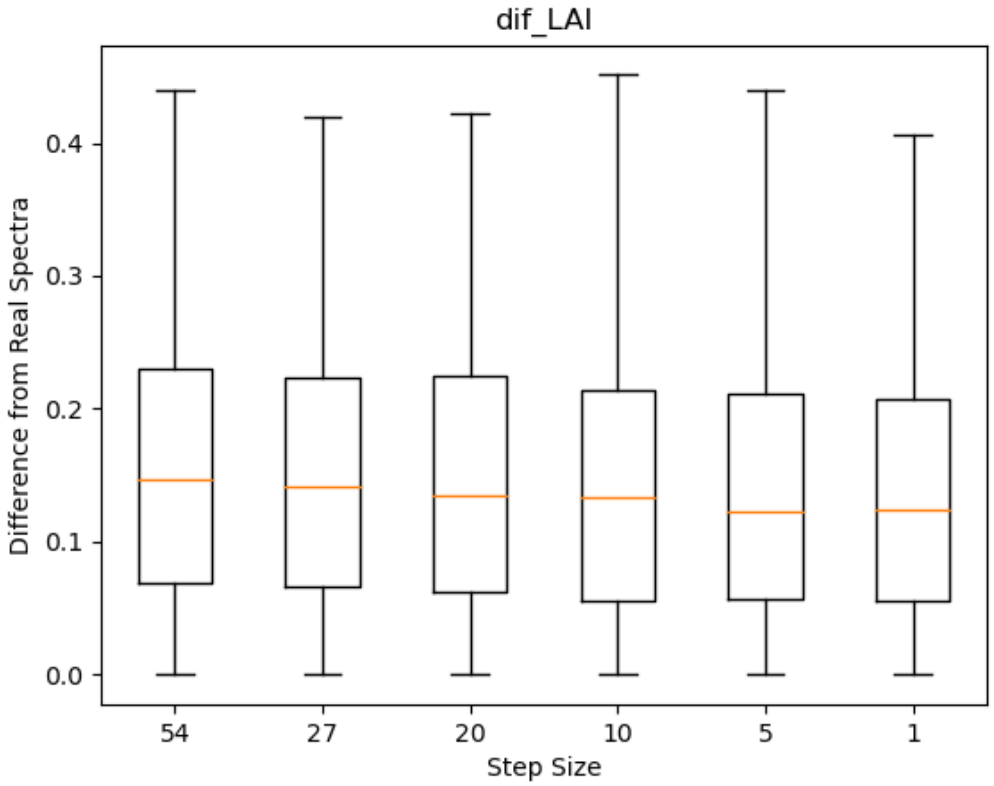

mSR705
Difference from Real Spectra
3.0
2.5
2.0
1.5
1.0
0.5
0.0
54
27
20
10
5
1
Step Size

1


nd_leaf_water
Difference from Real Spectra
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
54
27
20
10
5
1
Step Size

2


d_water_stress
Difference from Real Spectra
0.0
0.5
1.0
1.5
2.0
54
27
20
10
5
1
Step Size


dif_LAI
Difference from Real Spectra
0.0
0.1
0.2
0.3
0.4
54
27
20
10
5
1
Step Size