High Resolution Multispectral Image and Machine Learning Algorithms for Saltmarsh Classification and Biomass Modelling

Sikdar Mohammad Marnes Rasel

High Resolution Multispectral Image and Machine Learning Algorithms for Saltmarsh Classification and Biomass Modelling

By

Sikdar Mohammad Marnes Rasel

Department of Environmental Sciences
Faculty of Science and Engineering
Macquarie University
New South Wales, Australia

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

November 2017

Declaration

I, Sikdar Mohammad Marnes Rasel, declare that the thesis entitled “High Resolution Multispectral Image and Machine Learning Algorithms for Saltmarsh Classification and Biomass Modelling” contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Signature: Date:

Acknowledgements

First and foremost I would like to thank the Almighty Allah for his blessings, protection and guidance during the whole study period. My special thanks to iMQRES (International Macquarie University Research Excellence Scholarship) grant which provided me with a fellowship to study and join with people from different countries around the world. I am very much grateful to my employing organization 'Soil Resources Development Institute', Ministry of Agriculture, Bangladesh for providing the opportunity (Study leave) to study in abroad.

I am very much grateful to Dr.Michael Chang, my principal supervisor for motivating, guiding and supporting me through my PhD. My heartfelt thanks go to my co-supervisors Professor Neil Saintilan and Dr Tim Ralph for their continuous support, invaluable suggestions, constructive feedback and comments. Neil, I still remember the early days, when you helped me in my fieldwork, made a collaboration with The Commonwealth Scientific and Industrial Research Organization (CSIRO), Canberra for satellite image and other assistance.

Sincere thanks to Russell Field, Scientific Officer of the Department of the Environmental Sciences, for his support in field work, guidance and supervision in the laboratory analysis. I am deeply indebted to Daniel Sloane and Carl Helander who helped me in my fieldwork.

I have been fortunate to come across wonderful officemates, Shahriar Rahman, Provakaran, Mingzhu Wang, Alaa ElGendawy, Abdus Sattar and Ashraful Alam. I would like to thank them for their friendship and valuable time.

Last my everlasting gratitude goes to my parents, my wife and son, my brothers and sisters for their prayers, encouragement and moral support. You have always special place in my heart. This work is dedicated to my family.

Sikdar Mohammad Marnes Rasel
Sydney, Australia
November, 2017

Table of Contents

List of figures

List of tables

List of abbreviations

The following table describes the significance of various abbreviations and acronyms used throughout the thesis

Abbreviations	Meaning
ANN	Artificial Neural Network
CASI	Compact Airborne Spectral Imager
CART	Classification and Regression Tree
DTC	Decision Tree Classifier
EEC	Ecological Endangered Community
FLAASH	Fast Line-of-sight Atmospheric Analysis of Hypercube
HRSI	High Resolution Satellite Image
LiDAR	Light Detection and Ranging
MLC	Maximum Likelihood Classifier
MNF	Minimum noise Fraction
NIR	Near infraRed
OBIA	Object-Based Image Analysis
PCA	Principal Component Analysis
ROI	Region of Interest
REF	Recursive Feature Elimination
RF	Random Forest
SAM	Spectral Angle Mapper
SNR	Signal Noise Ratio
SVM	Support Vector Machine
SVR	Support Vector Regression
VSURF	Variable Selection Using Random Forest
VNIR	Very near Infra-Red

Executive summary

Saltmarsh is one of the important vegetation of wetlands. Due to a range of pressures, it has been declared as an Endangered Ecological Community (EEC) in Australia. Therefore, monitoring and mapping of the distribution of saltmarshes species are important to wetland management, conservation and distribution. This rigorous task requires intensive fieldwork and collection of ancillary data that are time-consuming. Remote Sensing offers a practical and economic means of plant sciences classification and biomass modelling. However, selecting suitable remote sensing systems and their data are important for mapping saltmarshes. Hyperspectral remote sensing can be used to monitor this endangered community. However, there are some crucial limitations of hyperspectral data that have been found in the current study area and discussed in the introduction. To overcome these limitations, Worldview-2 with its higher spatial (1.84 m) resolution is seen as a trade-off between the advantages of multispectral resolution satellite data and hyperspectral data.

In this thesis, imagery acquired by three different sensors were used to compare the performance of machine learning classification methods and biomass regression models. Maximum Likelihood Classifier (MLC) and two advanced algorithms, Random Forest (RF) and Support Vector Machine (SVM) were used for classification. These two algorithms were also tested to develop a biomass model for Sporobolus virginicus species using multispectral Worldview-2 data. Reflectance and NDVI based vegetation indices derived from 8 bands of Worldview-2 multispectral data were used for four experiments to develop the biomass model. Proportional reduction of sample size (100 % to 33%) was focused to test the potentiality of both algorithms. RF showed significant changes in overall accuracy when the sample size was reduced from 100 % to 33%. Conversely, there were no significant changes in the accuracy for SVM when the sample size equally dropped from 100 % to 33%. When biomass model for RF (R2 =0.72, RMSE = 0.166 kg/m2) and SVR (R2 = 0.66, RMSE = 0.200 kg/m2) were compared, there was a significant ($p =< 0.0001$) difference was observed. Further research is crucial to explore the sensitivity of this method with the spatial autocorrelation of the training samples for random forest application in saltmarsh monitoring.

Chapter 1

General Introduction

1.1 Saltmarsh and its importance

Wetlands are valuable natural resources due to their useful and viable economical products. At the same time, the whole wetland ecosystem or its subdivided parts are a considerable scientific interest due to their biological diversity, ecological function and process. Saltmarsh is one of them that is an intertidal community of plants, such as sedges, rushes, reeds, grasses, succulent herbs and low shrubs. Saltmarsh can tolerate high soil salinity and occasional inundation with salt water (Daly 2013). The term 'saltmarsh' is used to describe individual plants, groups of plants and the general estuarine habitat dominated by these plants. Saltmarshes tend to occupy the hyper-saline soils of the upper intertidal zone, where saltwater inundation occurs less frequently (usually only during high spring tides) (Daly 2013). For example, Smooth Cordgrass (*Spartina alterniflora*) and Salt coach grass (*Sporobolus virginicus*) are more effective at disposing of excess salt and can better tolerate having its roots submerged by the daily tides (Figure 1-1). Oppositely, Salt Marsh Hay (*Spartina patens*) and Common Reed (*Phragmites australis*) stay at higher levels of the marsh which are only exposed to the highest tides. These communities are generally found growing on the landward side of mangroves and are made up of salt tolerant, flowering plants in the form of low growing shrubs, herbs and grasses.

Up to mid-nineteenth century, this ecosystem was treated as boggy swamps and wastelands of little practical use. Due to this many saltmarsh areas have drained, turn in to a degraded form for human interference, reclaimed for other purposes or otherwise lost (Gedan et al. 2009). Ecological value and scientific concern for this ecosystem has raised over the past few decades and now treated as significant ecological communities that provide key habitat for other marine fauna(Adam 1993). It has been recorded this special ecosystem provides habitat and shelter for fish and a special support for some commercially and recreationally important species when it is inundated (Daly 2013). Saltmarsh, mangrove, and seagrass ecosystems all have relatively high rates of sediment carbon burial. Globally at least 430 Tg of carbon is stored in the upper 50 cm of tidal saltmarsh soils (Chmura 2009; Chmura et al. 2003). Saltmarsh provides an important buffer between land and reef, as they filter land runoff and improve the quality of water entering into the landward part. They also serve to buffer the coastline from storms and cyclones. They trap and stabilise sediment and dampen the effects of flood water (Daly 2013). Therefore, monitoring and mapping of saltmarsh species distribution, quality, and quantity assessment are an important consideration in wetland management. This rigorous task involves natural resources inventory, species mapping, habitat characterization and water quality assessment over a continuous period of time (Carpenter et al. 1999). In addition, it requires an up-to-date spatial information about the magnitude and the quality of vegetation cover in order to initiate vegetation protection and wetland restoration programme (He et al. 2005). However, wetland species discrimination mapping

requires intensive fieldwork, taxonomical information, and ancillary data, and the visual estimation of percentage cover for each species to be classified. This is time-consuming, cost-oriented and labour intensive and sometimes inapplicable due to the poor accessibility, and is thus, only practical on relatively small areas (Lee and Lunetta 1995). Due to these unique features and problems associated with the wetland ecosystem, mangroves and saltmarshes have been the priority for protection by the Ramsar Convention. The Ramsar Convention on Wetlands is the most famous and widely adopted wetlands protections projects. The Convention on Wetlands of International Importance holds the unique distinction of being the first modern treaty between nations aimed at conserving natural resources. The signing of the Convention on Wetlands took place in 1971 at the small Iranian town of Ramsar. Twelve NSW wetlands are listed under an international convention that aims to protect their ecological character. The Hunter Wetlands Centre Australia is a component of the Ramsar site, and is significant for a range of plant communities that have been successfully re-introduced to the site. That is why developing a dynamic and easy monitoring tool is very essential to monitor this unique ecosystem. Remote sensing offers a practical and economical means of species classification and estimate the biochemical and biophysical parameters of the wetland species and it can make field sampling more focused and efficient. For more than two decades, various remote sensing-based methodologies have been used to obtain the condition and extent of this ecosystems

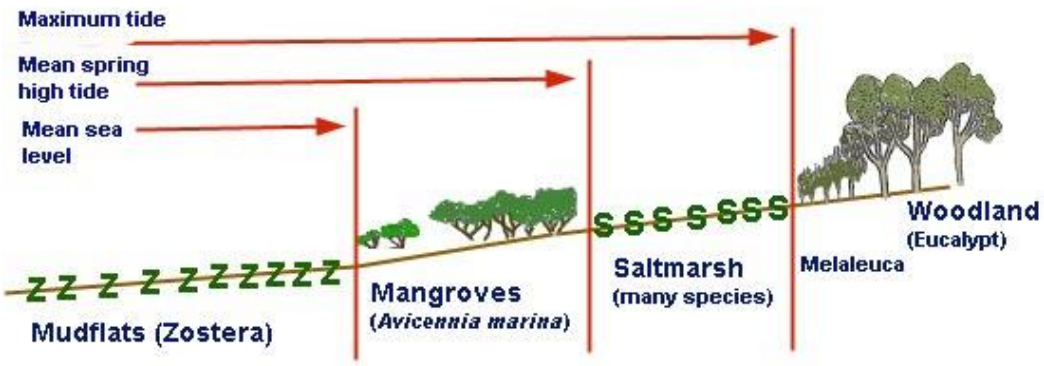

Figure 1-1: Thematic zonation of saltmarsh in respect of upland vegetation (i.e. Mangrove) and sea water (adapted from http://www.mesa.edu.au/saltmarsh/saltmarsh02.asp accessed date : 12 August 2018)

1.2 Remote sensing of wetland and saltmarsh ecosystems: Selection of accurate satellite image

Remote Sensing offers archive data for change detection over time for vegetation distribution in wetland. Its digital format data can be easily integrated into Geographic Information System (GIS) for more analysis (Ozesmi and Bauer 2002; Shaikh et al. 2001). Remote Sensing analysis have been useful for vegetation types, species classification (Carle et al. 2014) and biomass studies (Mutanga et al. 2012). It helps to determine which saltmarsh or other wetland species are spectrally separable and which bands and dates are best for species discrimination. For this advantage, many researchers have used multispectral data (i.e. SPOT imagery and Landsat TM) to classify general vegetation classes (Harvey and Hill 2001; Li et al. 2005). However, coarse to medium resolution image (Landsat TM and SPOT) data have been shown insufficient for saltmarsh and other wetland species classification due to the following limitation:

a) Broad nature of multispectral wavebands in respect to sharp ecological gradient with narrow vegetation units in saltmarsh ecosystem.
b) Spatial and spectral resolution are not high enough to detect vegetation class. Mostly per pixel classification method of multispectral image dominated by mixed pixel are often incapable of producing accurate species classification.

To overcome these limitations, many authors have used and recommended hyperspectral data to discriminate and map wetland vegetation at the species level(Rosso et al. 2005; Schmidt and Skidmore 2003). Even the use of hyperspectral techniques has been extended into measuring the biophysical and biochemical properties such as leaf area index (LAI), biomass, and water content of wetland vegetation (Proisy et al. 2007; Rendong and Jiyuan 2004). However, there are also some crucial limitations of airborne hyperspectral data that we experienced for our study area. First of all, the high cost to purchase aerial data if it is available. If data are unavailable for a specific study area, new acquisition is subject to season, sun illumination, weather conditions, flight schedules and aviation restrictions. In addition, time and processing costs of high dimensional (due to its narrow continuous bands) data are severe obstacles to general users, (e.g. wetland managers). In addition, the processing time and cost for hyperspectral data (Spaceborne or airborne) are too lengthy and expensive.

1.3 Remote sensing for wetland and saltmarsh ecosystem: Selection of the spatial unit and spectral separability analysis

Selection of spatial unit and sample size balance maintaining for advanced classifiers are still a controversial topic in image classification. It could be argued that classes with multimodal frequency distributions (i.e. agricultural land with different crop types and growing cycles) should have a greater number of samples in order for classes to be

accurately represented in the classifier than a spectrally and temporally well-defined class (i.e. waterbody) (Colditz 2015). Accordingly, advance image classification trees may sometimes suffer from a problem of unbalanced sample sizes problem. This means in the standard form of classification trees, the class with the highest number of samples determines the class label (Colditz 2015). Most of the algorithm compares pixel spectral with a reference or target. This reference spectra can be derived from the spectral library (for hyperspectral data), a region of interest within a spectral image or individual pixel within a spectral image (Ground truth data). Besides individual pixels, blocks of pixels (i.e. square arrays of pixels), and polygons are the spatial units commonly used for image accuracy assessment. Stehman and Czaplewski (1998) reviewed 33 map accuracy assessment to evaluate the choice of assessment unit and reported that 14 assessments used a pixel as the spatial unit, 10 assessments used a square block of pixels (e.g., 2x2, 3x3), and 9 assessments used a polygon. Although Richards (1996) and Strahler et al. (2006) support a pixel-based assessment, Congalton and Green (1999) recommended using a block of pixels or a polygon. So the choice and sett of spatial units (pixels, block or polygon) affect the accuracy report and associated with the sensors data. Hereby the selection of sensor for image classification is closely associated with the spatial resolution, a spatial unit of the reference sample and the size and quality of reference sample. For wetland ecosystem monitoring, the spatial unit is an excellent indicator for early signs of any physical and chemical degradation. Therefore, a spatial unit needs to be properly justified before its application for wetland monitoring.

1.4 Remote sensing for wetland and saltmarsh ecosystem: Selection of adequate algorithms for mapping and biomass modelling

There are different image classification methods used to analyse remote sensing data. Spectral Angle Mapper (SAM), Multiple Endmember Spectral Mixture Analysis (MESMA) and Spectral Mixture Analysis (SMA) are commonly used image classification methods. For multispectral image analysis, Maximum Likelihood Classifier (MLC) is a commonly used supervised classifier whereas ISODATA and K-Means are unsupervised methods (Richardson and LeDrew 2006; Srivastava et al. 2012). From the review of image analysis methods (Lu and Weng 2007), it may be considered that a single classification method cannot be treated as optimal. Based on specific objectives, sometimes it needs an advanced algorithm (ensemble) to produce accurate classification. Advanced classification algorithms include support vector machines (SVM), random forest (RF), artificial neural networks (ANN), and Decision tree classifier (DTC) (Adam et al. 2014; Adam et al. 2010; Petropoulos et al. 2012). The performance of these classifiers is varied with the quality of the remote sensing data either passive sensor (multispectral or hyperspectral data) or active sensor (LiDAR) data. One of the major problems relating to the supervised

classification lies in the definition of a proper training set size for an accurate learning of classifiers (Chi et al. 2008). Because the collection of ground-reference data is an expensive, time consuming and complex task. Therefore in many cases, the number of training samples is insufficient for a proper learning of classification systems. Within the supervised parametric classifiers, Maximum Likelihood Classification (MLC) is the most commonly used classifier and deliver excellent results when dealing with unimodal data. Because this classifier assumes a normal data distribution and has limitations when dealing with multi-modal input datasets (Liu et al. 2010). However, in most of the cases, normal distribution is the main violation of RS data because the nature and causes of spatial variation in images are not always easily understandable. In this regards some non-parametric supervised classifier, such as Support Vector Machine (SVM) (Mountrakis et al. 2011), Random Forest (RF)(Breiman 2001), Artificial Neural Network (ANN) (Mas and Flores 2008) and the Classification and Regression Tree (CART), are becoming increasingly popular classifiers for remote sensing data as they do not make any assumptions regarding frequency distribution. A very few studies (Adam et al. 2014; Zhang et al. 2015) are available where these algorithms have been used, and need further investigation in terms of accuracy, training sample size and quality, and parameters settings for the application of machine learning algorithms in a wetland environment.

1.5 Remote sensing for wetland and saltmarsh ecosystem: parameter optimization and variable selection

Among the advanced classifiers, SVM is appealing due to its impressive ability to successfully handle a small training dataset by producing higher classification accuracy. It is proposed as a superior classifier in remote sensing than the traditional methods like Maximum Likelihood Classifier (MLC) (Montero et al., 2005). MLC is a probabilistic algorithm and less suited when data are non-normally distributed. The MLC classifier assumes that reflectance values of each class are normally distributed. This is a common violation in remote-sensing data, especially when classes or even subclasses of the main class contain different spectral features (Kavzoglou & Reis, 2008). To overcome this problem, non-parametric classifiers such as decision tree classifier (DTC), artificial neural networks (ANN), and Support Vector Machine (SVM) are gaining priority in recent remote sensing classification (Kavzoglu and Reis 2008; Otukei and Blaschke 2010; Zhu and Blumberg 2002). Although there is ample proof of training sample size reduction for SVM, very few research are available for wetland ecosystem where data collection is really a challenge due to hazardous access. In addition, kernel selection for SVM and its relation with spectral separability for each species have not been explored yet for a complex environment, i.e. saltmarsh ecosystem. So still there is research opportunity to work with SVM parameters optimization and its relation with spectral separability analysis.

Thematic mapping and image classification are also associated with the features selection methods that are a crucial issue in machine learning (RF and SVM) dealing with applied classification and regression problems(Hastie et al. 2001). The potential benefits of feature selection methods are: facilitating data visualization and data understanding, reducing training and utilization times, and defying the curse of dimensionality (i.e. hyperspectral data) to improve prediction performance. Based on available literature, only a limited number of studies have been done, which explore advanced feature selection methods in multispectral data. Recently, Adam et al. (2014) applied these two algorithms on multispectral RapidEye data for a heterogenous landscape classification. Although RapidEye has five multispectral bands including Red-edge and NIR 1, there is a gap in knowledge of the performance of these algorithms on high-resolution Worldview-2 imagery. Worldview-2 (WV-2) can be spectrally differentiated from RapidEye as the former has three additional bands, called Coastal blue (Band1), Yellow (Band 4) and Near Infrared Band II (Band 8). So it is very important to evaluate the performance of all bands based on feature selection method using machine learning algorithms.

1.6 Remote sensing for wetland and saltmarsh ecosystem: Machine learning algorithms for saltmarsh biomass modelling

There is a crucial need to quantify large-scale plant productivity (i.e. above ground biomass) in coastal marshes for a better understanding of marsh resilience against sea level rise (Schile et al. 2014; Swanson et al. 2014). But this AGB estimation is labour intensive and not feasible at large spatial extents. To complement this, remotely sensed data are utilized to map vegetation types and provide better estimates of plant production (Goetz and Dubayah 2011). Since the 1980s there have been some successful application of remote sensing for saltmarsh biomass (Gross et al. 1987; Hardisky et al. 1983; Hardisky et al. 1984) using Normalized Difference Vegetation Index (NDVI). There are some major limitations of NDVI that have already been revealed by other studies (Gao et al. 2000;Tucker 1977) and recommended using narrow bands of hyperspectral data to overcome the limitations (Blackburn 1998; Thenkabail et al. 2000). However, there are some crucial limitations of hyperspectral data that we experienced for our study area and discussed in the introduction and first two chapters. To overcome these limitations, Worldview-2 with its higher spatial (1.84 m) and spectral (8 bands) resolution is seen as a tradeoff between the advantages of multispectral resolution satellite data and hyperspectral data (Mutanga et al. 2012; Rasel et al. 2016).

Another research area that is still challenging is to model biomass against remote sensing variables. Regression techniques are commonly used to relate remotely sensed information with biophysical variables but imitated to adequately capture the relationship and the spatiotemporal variability of the quantity (Kaheil et al. 2008). Moreover, multicollinearity is an important issue for multiple regression model especially when highly related variables (i.e bands of RS data) are selected as a predictor. An ensemble method, random forest (RF), has reduces the problem of the multicollinearity issue (Liaw and Wiener 2002b) and has been proved to reduce of bias and overfitting (Breiman 2001) of

a model. So there is still a huge scope for the application of machine learning algorithms to improve the accuracy for the prediction model of biomass estimation.

1.7 Overall problem statement

The first step of any image classification is to select the suitable image based on the condition of the study area, availability of the data and research budget. When data are available, it is also important to check the quality of the data. For example, EO-1 Hyperion hyperspectral sensor was launched in 2000 as a part of one-year technology validation/demonstration mission. However, this mission was decommissioned in January 2017. But within these 17 years, these hyperspectral data have been used in many areas including vegetation mapping. In this research, Signal Noise Ratio (SNR) of this data was tested and Hyperion data were compared with the Worldview-2 data to identify the correct spatial and spectral resolution that was necessary for our study area. The second important criteria for vegetation mapping and biomass modelling is to set the spatial unit (region of interest or pixel), that determine the homogenous and heterogeneous condition of the study area followed by the accuracy of the thematic map and prediction model. In this research, both of the spatial unit (ROI and pixel) have been tested to determine the right spatial unit for mapping accuracy followed by biomass model. The spatial unit is associated with the spatial analysis (like any classification) algorithms. Maximum likelihood (MLC) is the most common supervised classification method. However, in most of the cases normal distribution of the classes in the spectral domain is the main violation of remotely sensed data. Because sometimes the reflectance values of a main class and their subclass contain different spectral properties that makes the application of MLC more difficult and less accurate. In addition, adequate ground truth information and collection of sufficient samples of training and validation are also impractical for wetland ecosystem. Therefore, the application of machine learning algorithms are increasingly popular for land cover mapping. As machine learning algorithms are very new for wetland mapping and biomass estimation modelling, they need further investigations to improve accuracy and species-specific model for wetland ecosystem. Therefore, two advance machine learning methods, Random Forest (RF) and Support Vector Machine (SVM) have been tested for saltmarsh species classification and biomass prediction model. For a robust prediction model, important feature selection is also a criterion and a part of machine learning algorithm. There is ample proof for hyperspectral remote sensing for feature selection and dimensionality reduction (Pal 2006; Pal and Foody 2010). However, only a limited number of studies have been done, which explore advanced feature selection methods in multispectral data.

Variable selection for saltmarsh biomass estimates is challenging in terms of a single (i.e. water) and multi-modal distribution (seasonal variation of vegetation) features of

saltmarsh habitat. In this article, an embedded method recursive feature elimination (REF) was applied for Random Forest and SVM classifier. Another research area that is still challenging is to model biomass against remote sensing variables. Regression techniques are common to relate remotely sensed information (i.e. spectral bands or indices) with biophysical variables (i.e. biomass, leaf area index etc.). However, traditional regression models have limitations to adequately capture the relationship and the spatiotemporal variability of the quantity (Kaheil et al. 2008). Moreover, multicollinearity is an important issue for the multiple regression model, especially when highly related variables (i.e bands of RS data) are selected as a predictor. The multicollinearity problem can be alleviated using RF where a random subset of features is chosen for each tree (Cutler et al. 2007; Díaz-Uriarte and Alvarez de Andrés 2006; Liaw and Wiener 2002a). Therefore, two machine learning algorithms will be compared to develop a species-specific biomass model based on optimum parameters settings and selected features.

1.8 Research objectives

The primary aims of this thesis were to:

1. Investigate the signal noise ratio of spaceborne hyperspectral data to determine the importance of spatial and spectral resolution based on a fragmented saltmarsh ecosystem (Chapter 2).

2. Analyse spectral separability and determine the spatial unit (region of interest or pixels) suitable for saltmarsh species classification using Worldview -2 (1.84-meter spatial resolution) and Landsat 8 OLI (30-meter spatial resolution) data (Chapter 3).

3. Evaluate the potential of machine learning algorithms (Support Vector Machine and Random Forest) for saltmarsh classification based on pixel-based classification (Chapter 4).

4. Evaluate the application of a small training sample size and Support Vector Machine (SVM) for wetland saltmarsh classification (Chapter 5).

5. Investigate different variable selection approaches and equal allocation of training data to improve classification accuracy based on machine learning algorithms (Chapter 6).

6. Evaluate predictive models of aboveground biomass of a common saltmarsh species based on machine learning algorithms (Chapter 7).

1.9 Input data

Three different platforms satellite data have been used in this study. They are as follows:

Table 1-1. List of three different sensors used in this study

Data	Spatial resolution	Spectral Resolution
Spaceborne Hyperspectral data EO-1 Hyperion	30 meter	242 bands comprise of Very Near Infra-red (VNIR, 1-70) and Short Wave Infra-Red (SWIR, 71-242) bands.
Landsat 8 OLI (Operational Land Imager)	30 meter for band 1 to 7 and 9	Total 11 bands including two thermal and 1 panchromatic.
Worldview-2	1.84 meter	High Spatial Resolution multispectral data with a combination of eight spectral bands ranging from 400 nm to 1040 nm.

There are other satellite images like Sentinel that has the similar properties of Worldview-2. However, Sentinel-2 has 4 bands with a spatial resolution of 10 m and it has also three other bands with a spatial resolution of 20 m, whereas Worldview – 2 has eight bands with a spatial resolution 1.84 bands. Fragmented marsh patches are the main characteristics of our study area. That is why high spatial resolution (1.84m for WV-2) was the best choice for our study area. Another commonly used high-resolution satellite data is SPOT. SPOT-7 (launched in 2014) has four multispectral bands with spatial resolution 6.m. These four multispectral bands include Blue (0.455 µm – 0.525 µm), Green (0.530 µm – 0.590 µm), Red (0.625 µm – 0.695 µm) and Near-Infrared (0.760 µm – 0.890 µm). As one of the focus was to select high spatial resolution images, Worldview-2 was the best alternative among the available commercial satellite data.

Remote sensing and spatial analysis were performed in ENVI Classic, ERDAS and ArcGIS 10.3. All statistical calculation and analysis were performed in open source software R.

1.10 The study area

The study area as shown in Figure 1-2, is located in Tomago, Australia (Longitude 151°43'40.6" E to 151°46'19.4" E and latitude 32°47'21.9" S to 32°51'29.4" S) which is approximately 8 km south of Raymond Terrace and 10 km north of Newcastle on the east coast of Australia. The dominant tree species include *Avicennia marina* (Grey mangrove), *Casuarina glauca* (Swamp Oak or She-oak) and ground saltmarsh species includes *Sporobolus virginicus* (Salt couch) and *Phragmites australis* (Common reed). The study

area consists of a saline coastal wetland system comprising mangroves, saltmarshes and mixed mangrove-saltmarsh habitats (Rogers et al. 2014). About 4,257 ha of these wetlands are formally protected within the Hunter Wetlands National Park and this includes over 2,926 ha listed as internationally important under the Ramsar Convention (1984). Over the past decade, tidal reinstatement programs in the lower Hunter have restored tidal flow to several large wetland systems including Tomago (Winning and Saintilan 2009). Due to this, this study site is a focus of scientific research aimed at quantifying the benefits of wetland restoration. My interest in this study was to better understand the current status of theses degraded saltmarshes in relation to other associated habitats, such as mangrove and casuarina sp and identify the optimum mapping accuracy for each saltmarsh species from remote sensing data. In addition, some ancillary data collected from the Office of the Environment and Heritage (OEH) were also available for this study site.

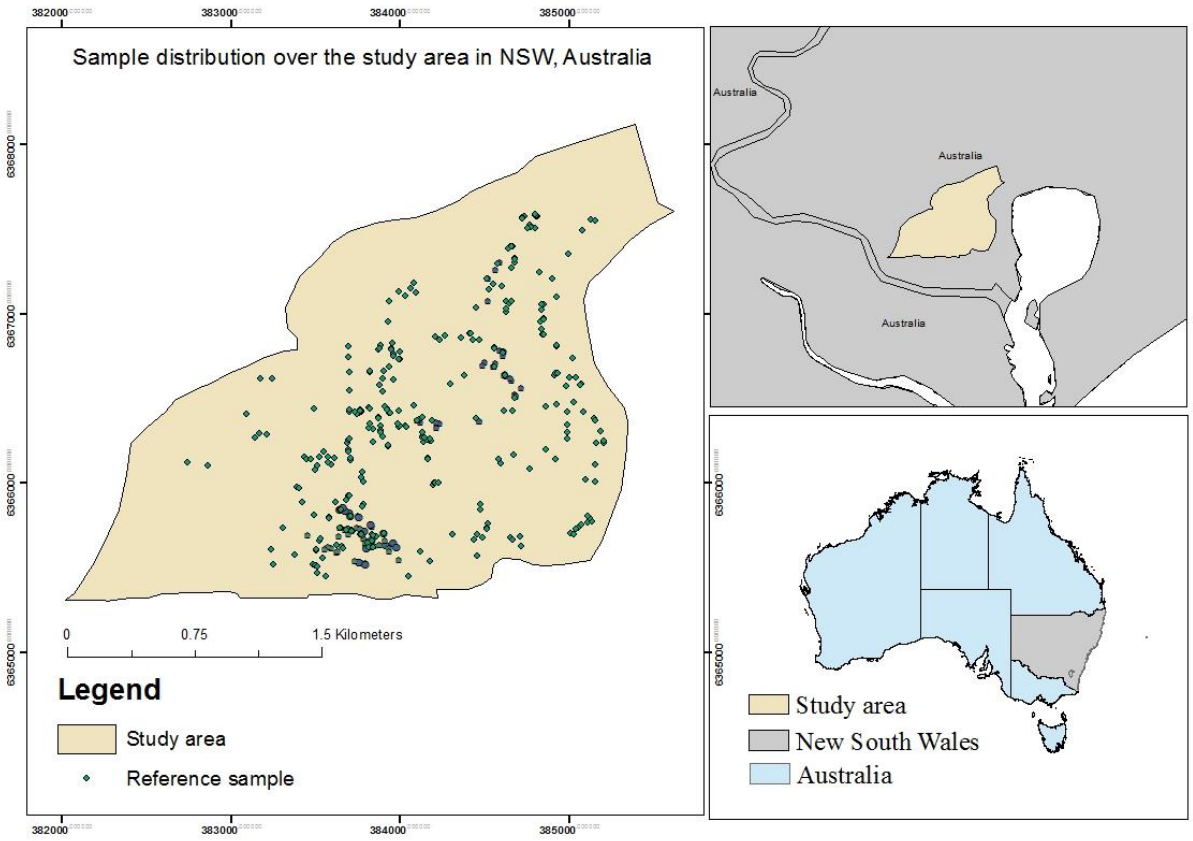

Figure 1-1: Australia country boundary and location of Wetlands (study site) in New South Wales (NSW) state. The first inset picture shows the Hunter Wetland National park in NSW (Clockwise), the second one shows the location of NSW state in Australia and the last and main view picture shows the spatial extent of the study area.

1.11 Land cover and the species classes

Land cover type of this area can be divided into three distinct strata (Table 1-2). Details explanation of these are:

Table 1-2: Land cover strata and species types of the study area.

Land cover	Specific Classes (n= 8)	Description
Trees	Grey Mangrove (*Avicennia marina (*Forssk.))	Known as grey mangrove or white mangrove, is a species of mangrove tree classified in the plant family Acanthaceae.
	Swamp Oak or She-oak (*Casuarina glauca)(* Sieb. Ex Spreng)	Commonly known as the swamp she-oak, swamp oak, grey oak, or river oak. It is a *Casuarina species* native to the east coast of Australia. It is found from central Queensland coast south to the southern New South Wales coast.
	Die-back Swamp Oak	Die-back in *Casuarina glauca* has occurred in some parts of the study area.
Saltmarsh and Grass	*Phragmites australis* (Cav.) Steud.	This common reed forms large beds in shallow water. It has round, hollow stems, which typically grow to 2m in height
	Sporobolus virginicus (L) Kunth	*Sporobolus virginicus*, known by numerous common names including seashore dropseed, marine couch, sand couch and, salt couch grass, is a species of grass with a height of 10 to 50 cm. This is the dominant saltmarsh species in study area.
	Perennial Grass	This group consists of common grass species found in the study area, that is distinct from the other saltmarsh classes.
Water and Wetland	Water	Shallow to deep water within the wetlands areas of the study site.
	Wetland soil	Muddy bare land within the wetland, most of the time it is inundated with tidal water.

1.12 Methodological flowchart and data selection

Methodological flow, the main component of each chapter and organization of chapters are presented in the following flowchart.

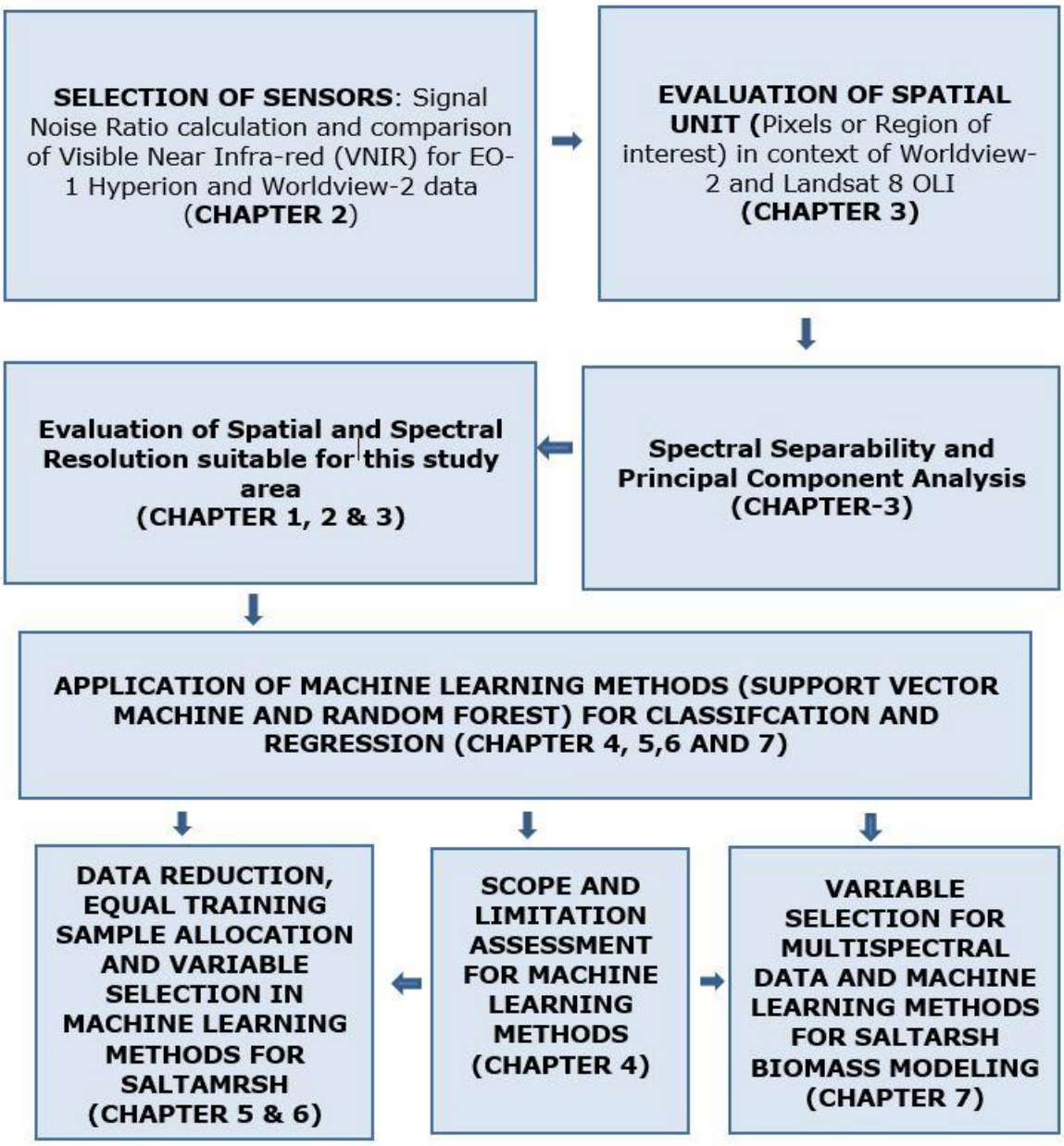

Figure 1-2: Orientation of six (six) main chapters and their link to each other

1.13 Thesis outline

This thesis comprises of an entire eight chapters, including a general introduction, six core chapters, and a synthesis. Each core chapter has been provided as a standalone research paper that has been published (Table 1-3) in or submitted to the peer-review ISI journals. The structure of the chapters is as follows.

Chapter 1 presents the research background, general objectives, and also describes the study area and outline of the thesis.

Chapter 2 deals with the effect of Signal Noise Ratio (SNR) in the visible-near infrared domain of EO-1 Hyperion data and its comparison with Worldview-2 for saltmarsh ecosystem. It also highlights the endmember identification from EO-1 Hyperion data.

Chapter 3 analyses spectral separability and Principal Component Analysis (PCA) of saltmarsh species from Worldview 2 and Landsat 8 OLI data using a block of pixels or region of interest (ROI) based training sampling. This chapter also highlights the application of traditional supervised classification, Maximum Likelihood Classifier (MLC).

Chapter 4 synthesizes the importance of machine learning classifiers (Support Vector Machine and Random Forest) and their application for saltmarsh classification and biomass estimation. This is a literature review based on the application of SVM and RF for classification and biomass modelling. Special emphasis was given to the saltmarsh ecosystem.

Chapter 5 applies Support Vector Machine (SVM) with small training samples for wetland saltmarsh environment. It applies parameter optimization, kernel selection and its relation spectral separability analysis.

Chapter 6 applies Support Vector Machine (SVM) and Random Forest (RF) for feature selection and classification for saltmarsh ecosystem based on pixel-based training sampling. This chapter also focuses on the 'equal training sample allocation' for SVM and RF. This chapter uses grid search methods for parameter optimization and uses recursive feature elimination method for variable selection.

Chapter 7 evaluates the performance of Support Vector Machine (SVM) and Random Forest (RF) for biomass estimation of a specific saltmarsh species, *Sporobolus virginicus*. This chapter also investigates feature selection algorithms from vegetation indices and spectral bands from Worldview-2 data

Chapter 8 provides an overview of the significant research findings of the thesis. The contributions of the undertaken studies in this thesis in the concept of Ecologically Endangered Community (EEC) are discussed. Further research direction based on remote sensing data for saltmarsh ecosystem are also proposed.

Table 1-3: Current status of research outputs (publication) originated from this thesis.

	Description	Status
Chapter 1	Title: General Introduction	
Chapter 2:	**Title : Signal Noise Ratio calculation and Selection of the sensor for saltmarshes monitoring**	
	Conference article-1: Endmember identification from EO-1 Hyperion L1_R Hyperspectral data to build saltmarsh spectral library in Hunter Wetland, NSW, Australia. Paper presented at the SPIE Remote Sensing. Proc. SPIE 9637, Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII, 96371O (October 14, 2015); doi:10.1117/12.2195444.	**Published**
	Peer-reviewed article -1: Comparison of very near infrared (VNIR) wavelength from EO-1 Hyperion and Worldview 2 images for saltmarsh classification. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., III-8, 85-92. doi:10.5194/isprs-annals-III-8-85-2016.	**Published (*Peer reviewed*)**
Chapter -3:	**Title: Spectral Separability analysis of saltmarsh species based on Worldview-2 and Landsat 8 OLI data**	
	Conference article-2: Comparative analysis of Worldview-2 and Landsat 8 for coastal saltmarsh mapping accuracy assessment. Proc. SPIE 9864, Sensing for Agriculture and Food Quality and Safety VIII, 986409 (May 26, 2016); doi:10.1117/12.2222960.	**Published**
Chapter-4:	**Title: Scope for saltmarsh classification and biomass estimation using Support Vector Machine (SVM) and Random Forest (RF): A review**	
	Peer-reviewed article-2: Scope for saltmarsh classification and biomass estimation using Support Vector Machine (SVM) and Random Forest (RF): A review	**Submitted** to the Estuarine, Coastal and Shelf Science
Chapter-5:	**Title: Support Vector Machine (SVM) classifier with small training samples for wetland saltmarsh environment**	
	Peer-reviewed article-3: Support Vector Machine (SVM) classifier with small training samples for wetland saltmarsh environment	**Submitted** to Geocarto International
Chapter-6:	**Title: Application of Machine learning algorithms (Random Forest and Support Vector Machine) for Saltmarsh classification**	

	Peer-reviewed article -4: An evaluation of equal training sample allocation for a saltmarsh environment using Random Forest and Support Vector Machine (SVM) classification algorithms.	**Submitted** to to the Estuarine, Coastal and Shelf Science
Chapter-7:	**Title : Saltmarsh biomass modelling using Random Forest (RF) and Support Vector Machine (SVM) regressions from multispectral data**	
	Peer-reviewed article-5: Saltmarshes biomass modelling using Random Forest (RF) and Support Vector Machine (SVM) regressions from multispectral data	**Revised version submitted** to the International Journal of Digital Earth
Chapter-8:	Synthesis : High spatial resolution multispectral image and machine learning algorithms for saltmarsh classification and biomass modelling	

Chapter 2
Endmember identification from EO-1 Hyperion and a comparison between hyperspectral and multispectral images for a saltmarsh community in the Hunter Wetland National Park

This chapter is based on:

Rasel, S. M., Chang, H. C., Diti, I. J., Ralph, T., & Saintilan, N. (2016). Comparison of very near infrared (VNIR) wavelength from EO-1 Hyperion and Worldview 2 images for saltmarsh classification. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 3(8); doi :10.5194/isprs-annals-III-8-85-2016.

Sikdar M. M. Rasel ; Hsing-Chung Chang ; Tim Ralph ; Neil Saintilan; Endmember identification from EO-1 Hyperion L1_R hyperspectral data to build saltmarsh spectral library in Hunter Wetland, NSW, Australia. Proc. SPIE 9637, Remote Sensing for Agriculture, Ecosystems, and Hydrology XVII, 96371O (October 14, 2015); doi:10.1117/12.2195444.

Abstract

Saltmarsh is one of the important communities of wetlands. Due to a range of pressures, it has been declared as an EEC (Ecological Endangered Community) in Australia. In order to correctly identify different saltmarsh species, development of distinct spectral characteristics is essential to monitor this EEC. This research was conducted to classify saltmarsh species based on spectral characteristics in the VNIR wavelength of Hyperion Hyperspectral and Worldview-2 multispectral remote sensing data. Signal Noise Ratio (SNR) and Principal Component Analysis (PCA) were applied in Hyperion data to test data quality and to reduce data dimensionality respectively. FLAASH atmospheric correction was done to get surface reflectance data. Based on the spectral and spatial information a supervised classification followed by Mapping Accuracy (%) was used to assess the classification result. SNR of Hyperion data was varied according to season and wavelength and it was higher for all land cover in VNIR wavelength. There was a significant difference between radiance reflectance spectra. It was found that atmospheric correction improves the spectral information. Based on the PCA of 56 VNIR band of Hyperion, it was found that 16 PCs contain 99.83 % variability. Later 16 bands were compared with 8 bands of Worldview-2 for classification accuracy. Overall Accuracy (OA) % for Worldview-2 was increased from 72 to 79 while for Hyperion, it increased from 70.47 to 71.66 when bands were added orderly. Considering the significance test with z values and kappa statistics at 95% confidence level, Worldview-2 classification accuracy was higher than Hyperion data. Based on the small patch size and vegetation distribution pattern high spatial resolution data was more effective compared to narrowband coarse spatial resolution data. As ground truth data collection is a difficult task for wetland ecosystem, spectral separability analysis from pixel and region of interest (ROI) based sampling may be considered in future work to save time and cost.

Keywords : EO-1 Hyperion, Signal Noise Ratio, FLAASH, Worldview -2, PCA.

2.1 Introduction

Advanced remote sensing technology like hyperspectral data, with an ability to monitor more detailed changes in vegetation and species composition (Zomer et al. 2009) will expand opportunities for saltmarsh monitoring and mapping. High spatial and spectral resolution remote sensing data with more advanced geospatial technology allows mapping any changes in vegetation cover using species signature analysis. Some authors used airborne hyperspectral data, particularly, Compact Airborne Spectral Imager (CASI) imagery for mapping and monitoring salt marshes (Belluco et al. 2006; Hunter and Power 2002; Thomson et al. 2003), still the data acquisition is a time-consuming and expensive activity for airborne hyperspectral data(Hunter and Power 2002). This problem that we had experienced for our study area has been discussed in the general introduction chapter. In this circumstances, narrow band (198 calibrated bands) coarse spatial resolution EO-1 Hyperion data might be an alternative. But it has a low signal to noise ratio in comparison to airborne hyperspectral sensors. The result of the signal in this spacecraft lost to atmospheric absorption and the reduced energy available from surface reflectance at orbital altitude. Moreover, detector arrays used in this sensor were "spares" originally designed for another purpose, which further decreases the signal to noise ratio (Jupp and Datt 2004). On the other hand, High-Resolution Satellite Imagery (HRSI) data products are routinely evaluated during the so-called in-orbit test period, in order to verify if their quality (SNR and other radiometric properties) fits the desired features. High-resolution satellite data and its recent advances have significantly improved the coastal and saltmarsh vegetation mapping. Due to the sub-meter spatial resolution and the advantage of satellite platform for repeated data acquisition with the minimal coast, Space Imagines' IKONOS and Digital Globe's Quickbird-2 has facilitated the routine change detection monitoring of both salt-marsh and terrestrial vegetation. For example, with high-spatial resolution QuickBird-2 satellite remote-sensing data Wang et al. (2007) mapped both terrestrial and submerged aquatic vegetation communities of the National Seashore Suffolk County, New York, and achieved approximately 82% overall classification accuracy for terrestrial and 75% overall classification accuracy for submerged aquatic vegetation and provided an updated vegetation inventory and change analysis results. In another study, Ouyang et al. (2011) used Quickbird imagery to efficiently discriminate salt-marsh monospecific vegetation stands using object-based image analysis (OBIA) classification methods in terms of accuracy than pixel-based classification method.

Considering the prospect of HRSI, no mentionable research has been done using Worldview-2 for saltmarsh classification although it has finer spatial and spectral resolution in compare to Quickbird. Moreover, we added high spectral resolution Hyperion data to test the efficiency of spectral and spatial resolution. As the Signal-Noise-Ratio (SNR) is one of the important properties of Hyperion data, we considered this property to test the

quality of the data before selecting the spectral wavelength for comparison. We also considered the information redundancy of Hyperion data. Within 242 original bands, the information content of the one band can be fully or partially predicted from the other band in the data. This redundancy exists due to the high correlation between bands, especially between adjacent bands (Jiang et al. 2004). Hence specific algorithms like Principal Component Analysis (PCA), Minimum Noise Fraction (MNF) are generally used to remove redundant dimensions and to select optimum bands number for further analysis.

The current study explores how spectral resolution (Very Near Infrared) part and spatial resolutions of satellite images affect salt-marsh vegetation classification. For saltmarsh monitoring and management, it is essential to have a knowledge of the spatial distribution of salt-marsh vegetation types. This study focuses on the potentiality of high-spatial and high-spectral resolution satellite data for reliably salt-marsh vegetation species classification with the help of extensive ground truth data. The objectives were (1) to segregate effective number of bands from Hyperion data to minimize redundancy of information; (2) to evaluate data quality based on Signal Noise Ratio (SNR) (3) to identify the efficiency of Visible to VNIR wavelength for saltmarsh classification from two sensors and (4) to assess the efficiency of high spatial resolution in context of coarse spectral resolution of the classes of interest. Although this study has ignored Short-Wave Infrared (SWIR) part from Hyperion, the results can then be used as a baseline information for further saltmarsh related monitoring program where spatial resolution is a fact due to a small patch of species distribution.

2.2 Materials and methods

2.2.1. *Remote sensing and other ancillary data*

Satellite imagery from two sensors was used for this research. High-spectral resolution EO-1 Hyperion data and high-spatial-resolution data from Worldview-2 and were used to compare the sensor capabilities in discriminating salt-marsh vegetation. Worldview-2 images have 0.46 m pixel resolution in the panchromatic mode and 1.84 m resolution in the multispectral mode. The multispectral mode consists of eight broad bands in the coastal blue (400-450 nm), blue (450–510 nm), green (510-580 nm), Yellow (585- 625 nm), red (630-690 nm), red edge (705 – 745), NIR1 (770-895) and NIR2 (860-1040) parts of the electromagnetic spectrum. EO-1Hyperion images have 242 narrow bands and a pixel resolution of 30 m. The Worldview-2 satellite data were captured on 5th May 2015, and the EO-1 Hyperion satellite data were captured on 6th June 2015.

2.2.2. *Field data*

For ground truth, an extensive fieldwork was conducted in the study area on 10th to 12th June 2015. The stratified sampling design was followed based on mangrove, saltmarsh and other cover types (water and grass etc.). Although homogeneity was a crucial issue for sampling size, however each of the sample sites were at least 30 m × 30 m so that

the data collected could be used for the Hyperion as well as the Worldview-2 image training and classification. Sampling data included vegetation species class, percentage occurrence of each species within the selected plot and their global positioning system (GPS) locations. Total 156 sampling points and related information were recorded and divided into two parts for training and validation. 70% of samples were used to train data and rest 30% were used to validate the training result. Both mages were rectified using local council (https://maps.six.nsw.gov.au/) ground control points (GCPs) to WGS 84 UTM Zone 56 S projection system. The image-processing task was carried out in ENVI Classic, ERDAS IMAGINE 2015 and ArcGIS 10.2.

2.2.3. *Processing of EO-1 Hyperion data*

With a single scene of Hyperion observation for classification with training data, it is not necessary to use atmospherically correct image data (Datt et al. 2003). Because it tends to amplify noise levels and reduces the Signal Noise Ratio (SNR). Considering our scene of wetland ecosystem we did an atmospheric correction (Figure 2-1.) but calculated SNR well ahead with radiance data to assess data quality.

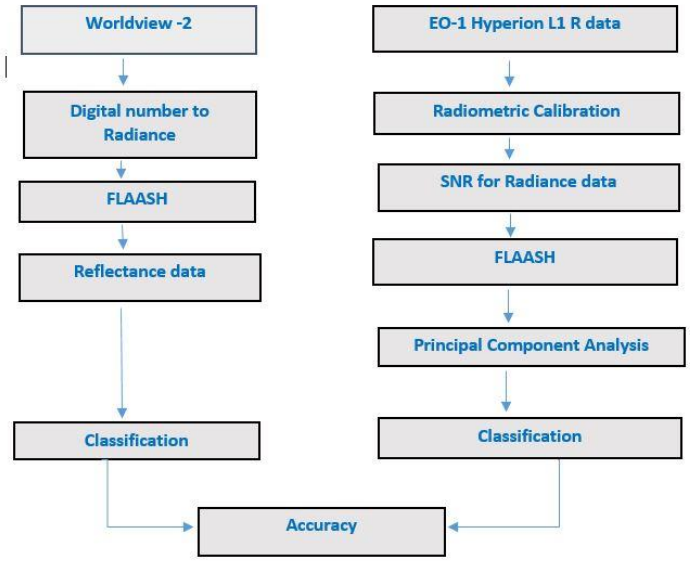

Figure 2-1: Processing flow of EO-1 Hyperion and Worldview-2 data

Elimination of bad bands based on reference and other information

Among the 242 bands of L1_R Hyperion data, it has been found that some bands are set zero during level-1 processing. They are bands from 1 to 7, bands from 58 to 76 and bands

from 225 to 242. The remaining 198 calibrated bands (Beck 2003) have been used for SNR calculation. Bands 77 and 78 were removed due to low SNR (Datt et al. 2003). Water absorption bands 120 – 122, 126 -132, 165-182, 185- 187 and 221 – 224 also removed (Beck 2003) but bands 123-125 have been retained in the image because some atmospheric correction programs like ENVI FLAASH require bands centres near 1380 nm in the strong water vapour wavelength for masking clouds. Thus 158 remaining calibrated bands used for radiometric calibration for radiance data followed by an atmospheric correction.

Correction of Stripping

Hyperion uses a push broom sensor where a poorly calibrated detector in either VNIR or SWIR has a slightly unbalanced responsivity from that of its neighbours or from its normal condition. It leaves a high-frequency error in the corresponding band of the image data known as "vertical stripes". Ultimately vertical stripes are an arrangement of 'abnormal pixel' either continuous or intermittent form. This abnormality of pixels rises due to detector failure, errors during data transfer, and improper data correction (Han et al. 2002) . Hyperion images also suffer from sensor optical properties which distort spectrograms (Scheffler and Karrasch 2013) commonly known as 'SMILE' effect. Smile effect is a shift of wavelength in the spectral domain (Yokoya et al. 2010). From different literature and study it is proved that uncorrected stripping effect will lead to faulty interpretation results of the data (Scheffler and Karrasch 2013) and it becomes very important when any research project dealing with a high similarity of spectral signature. As the current study deal with the spectral signature of similar saltmarsh vegetation, removal of the image stripes with the preservation of the original spectral information was very important. Based on the different existing de-stripping algorithm, "ENVI SPEAR – Tools – Vertical Stripe Removal" has been selected to a de-stripe subset of data. This method recalculates nearly the whole image information based on a global algorithm. The advantages of this global algorithm are that it removes low-frequency ('SMILE') effect from data (Yokoya et al. 2010).

Atmospheric correction and surface reflectance

The Hyperion data has been acquired in June 2015 when there is a spectral variation in trees and shrubs was expected. FLAASH is an atmopheric correction tool that corrects wavelengths in the visible through near-infrared and shortwave infrared regions, up to 3 μm (Somdatta and Chakrabarti 2010). The DN values of Hyperion L1_R data are scaled at-sensor radiance and stored as 16 bit signed integer that needs a radiometric calibration to get the absolute radiance. Then FLAASH atmospheric correction module has been selected to convert absolute radiance values in the image to its reflectance values (Solution 2016).

MNF transformation and Pixel Purity Index (PPI)

The MNF transform, as modified from Green et al. (1988) is implemented in ENVI. MNF transformation was done on de-stripped atmospherically corrected data to omit the bands those are highly correlated in terms of information. Whereas Pixel Purity Image means that each pixel value corresponds to the number of times that pixel was recorded as extreme. The general purpose of the PPI image is to associate spatial information (pixel locations) with the probability that each pixel represents a pure image endmember. From 32 MNF bands, first 10 bands (mostly uncorrelated) were used in the Purity Pixel Index (PPI) to extract pure pixel.

n-D visualization

The n-D visualizer can help to visualize the shape of a data cloud, with the image bands as plot axes (Solution 2016). This algorithm is commonly used to examine the distribution of points in n-D space in order to select spectral endmembers from an image (pixels that are pure, containing a unique type of material). This algorithm has been used in this study to identify endmember based on pure pixels.

2.2.4. *Processing of Worldview-2 data*

Commercials Worldview-2 products are delivered to the user as radiometrically corrected image pixels where the pixel values are calculated as a function of the amount of spectral radiance that enters the telescope aperture and the instrument conversion of that radiation into a digital signal. So it is very important to convert the digital number into radiance (figure 2-1) and then reflectance if we want to compare Worldview-2 data with other sensors that are related to spectral information.

Conversion of DN to radiance

The raw digital number (DN) has been converted to radiance data by applying the ENVI Worldview 2 calibration utility, available in ENVI v4.6 and greater. It uses the factors from the Worldview-2 metadata and applies the appropriate gains and offsets in order to convert those values to apparent radiance.

Atmospheric correction and surface reflectance

It is imperative that multispectral data be converted into reflectance prior to performing any spectral analysis. Currently, we have top atmospheric radiance data and have to be transformed into surface reflectance data. We used the FLAASH atmospheric module in

ENVI classic to remove atmospheric haze and to get surface reflectance data. FLAASH Module multiplies the reflectance data by 10000 to convert the heavy float type data (with decimals) into integers for fast calculations and lower data size. So, the output reflectance may exceed 10000 (also it include negative values related to shady areas within the image where FLAASH cannot calculate the solar irradiance at it. Here we used logical equations (Elsaid et al. 2014) to limit the reflectance data between 0 and 1 which is more reliable and comparable with most spectral libraries data range. Because
Surface reflectance = Surface radiance / Sun irradiance
So, surface reflectance should be less than 1.

2.2.5. *Selective bands from both sensors*

A spectral subset 56 bands have been selected (table 2.1) from 158 bands based on VNIR to SWIR wavelength (436.99 to 1043.59 nm) to match wavelength with HRSI Worldview-2 data. Hyperion is a narrow band hyperspectral data and contains redundancy of information within a narrow interval of wavelength. Minimum Noise Fraction was done to produce uncorrelated bands, segregate noise components and to reduce data dimensionality of 56 bands (Table 2.1).

Table 2-1: Selection of bands from two sensors

Worldview-2			EO-1 Hyperion	
Bands	Lower edges (nm)	Upper edges (nm)	Bands	Wavelength (nm) range
Coastal	400	450	B9-10	436.99 – 447.17
Blue	450	510	B11-16	457.34 – 508.22
Green	510	580	B17-23	518.39 – 579.45
Yellow	585	625	B24-B27	589.62 – 620.15
Red	630	690	B28-B34	630.32 – 691.37
Red Edge	705	745	B35-B41	701.55 – 762.60
NIR1	770	895	B42- B53	772.78-884.70
NIR2	860	1040	B64-B90	996.63-1043.59
8 Bands			56 Bands	

2.2.6. *SNR Calculation of EO-1 Hyperion data*

EO-1 Hyperion was designed for a one-year life as a test basis. But the instrument has continued to function well beyond two years with no degradation (Pearlman et al. 2003). It has already more than 10 years have passed and still, a lot of research are in progress with this sensors. So in our work, we tested SNR of 158 selected bands. After that 158

bands were radiometrically corrected in ENVI based on metadata information to get radiance data. These radiance data were used for SNR calculation. There are many analytical approaches (Atkinson et al. 2007) to calculate SNR. The simplest way is the mean over standard deviation method by which the SNR is expressed as the ratio of the mean signal over the standard deviation of a target interest. Standard approach uses a 50% albedo target, however user-defined targets based on interest can be selected to calculate SNR. Here SNR was calculated based on different season and different year of acquisition to find a relation with Hyperion proposed SNR.

2.2.7. *Classification algorithm and accuracy*

For supervised classification, the standard statistics "Maximum Likelihood Classifier" (MLC) algorithm was used. Overall Accuracy (OA), Producer Accuracy (PA) and User Accuracy (UA) were calculated based on the confusion matrix. For the accuracy of different vegetation classes, Mapping Accuracy percentage (MA %) was calculated based on the following equation (Congalton and Green 2008),

$$MA\ (\%) = (Pixels_{Correctly\ Classified}) / (Pixels_{Correctly\ classified} + Pixels_{Omissions} + Pixels_{Commissions}) * 100 \quad (2.1).$$

Where

Pixels $_{omissions}$ are the number of pixels assigned to other classes along the row of the confusion matrix relevant to the class considered.

Pixels commissions are the number of pixels assigned to other classes along the column of the confusion matrix relevant to the class considered.

2.3. Results

2.3.1 Results of the pre-processing of the Hyperion data

Effect of De-stripping on MNF transformation

Band selection and effect of de-striping on the selected band of Hyperion data were tested using the MNF transformation. The appearance of brightness gradient in MNF space or by in image difference technique is the symptom of 'smile' effect (figure 2-2.a, c) that appeared due to the push broom effect of the sensor. So removal of gradient either completely or partially is the indication of smile correction. As the global de-striping (ENVI Spear-tools vertical removal) is applied to this image, the result of applying the MNF transformation indicates that both the broad low-frequency effect (Bright gradient) (figure

2-2. a, c) and local stripes (figure. 2-2. e) have been removed significantly (figure 2-2 b, d, f). As seen from data image is clean.

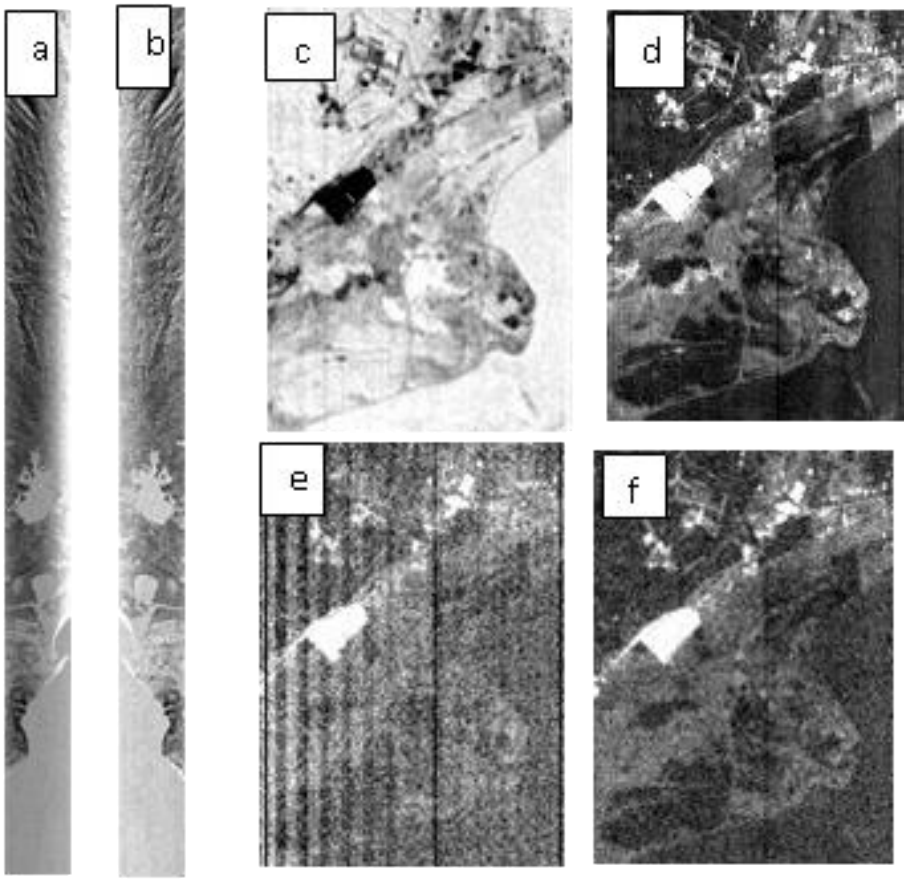

Figure 2-2. (a), (b) indicate the whole strip of Hyperion data for that study area before and after de-stripping respectively. From the image, the presence (a, c) of brightness gradient is the indication of smile and absence (b, d) is the indication of removal of smile. Image (e) and (f) represent result of global de-stripping algorithm before (e) and after (f).

Table 2-2: : MNF application to reduce the dimensionality of data.

MNF band	Eigenvalue	Percentage Variability	Cumulative Percentage
1	11.9191	13.2744	13.2744
2	9.0899	10.1235	23.3979
3	6.6892	7.4498	30.8478
32	1.0479	1.1671	80.7723

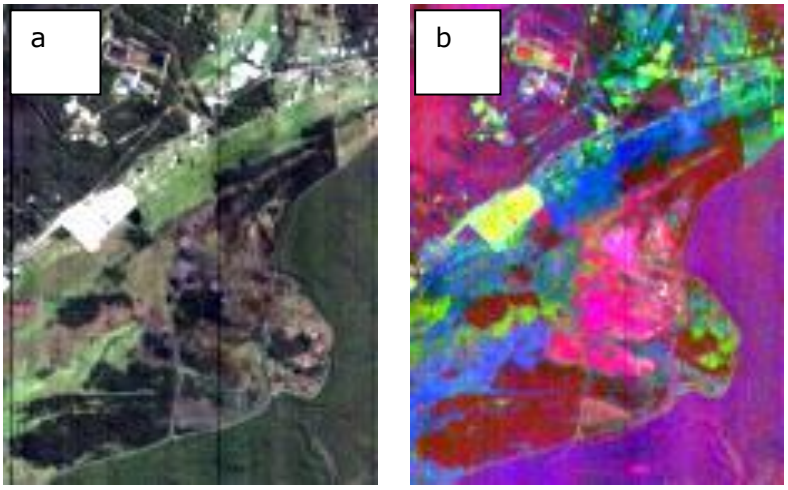

Figure 2-3: Study site (a) A true colour composite of band 29 (641nm), 23 (580nm) and 16 (509nm) (b) First three MNF band displays more colourful image and all land cover from EO-1 Hyperion.

Pixel Purity Index (PPI)

First 10 MNF bands were used in the Purity Pixel Index (PPI) to extract pure pixel. Here field training pixels were used to determine the threshold level. Based on the threshold value PPI algorithm was run with 500 iterations to select the pure pixels from the image. Based on the pure pixels cluster of n –D visualization different endmember were classified. Spectral features were explored in n-D visualization spectral plot based on the surface reflectance data of an atmospheric corrected image, where reflectance value was expressed as a scaled value (reflectance *10,000). In the study area, there is water and shady place that have a very low radiance. Those pixels may not model well in FLAASH and return negative reflectance. Sometimes these values appear very large but this is because the output from FLAASH is scaled by 10,000. In the study site, water spectra are often confusing with shade. From the spectral features, water can be identified with its very low reflectance from visible to SWIR part. High reflectance in visible to SWIR is the indication of built-up area. Vegetation group was characterized by typical vegetation absorption features due to photosynthetic pigments and water (figure 2-4).

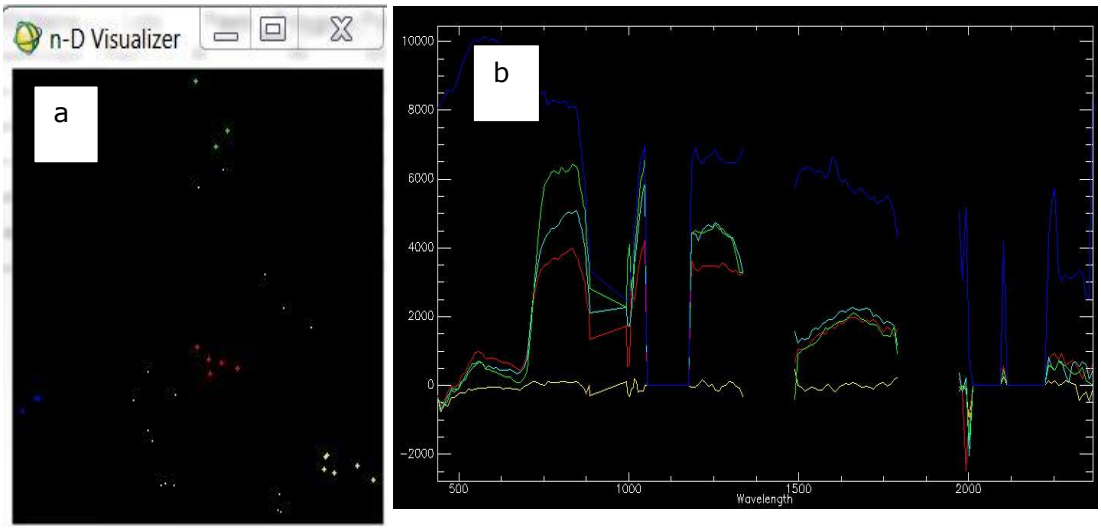

Figure 2-4: (a) Indicate group of pure pixel extraction by n-D visualization and (b) indicate respective endmember from each pixel group (recognized by separate colour).

2.3.2 Signal Noise Ratio Calculation (SNR)

SNR varies from 0 to 110 based on the season and acquisition time. SNR is highest in VNIR region for both dataset and ranges 0 to 40 with a maximum of 110 at 500 nm. Figure 2-5 and figure 2-6 show the estimated SNR for the study area in two different seasons.

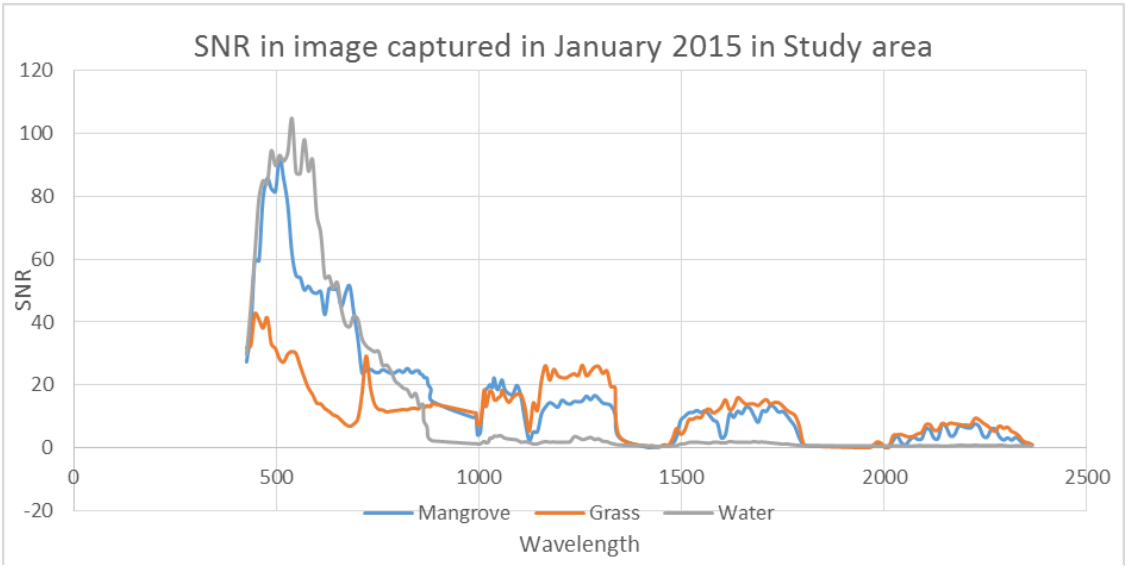

Figure 2-5: SNR of EO-1 Hyperion data (summer season).

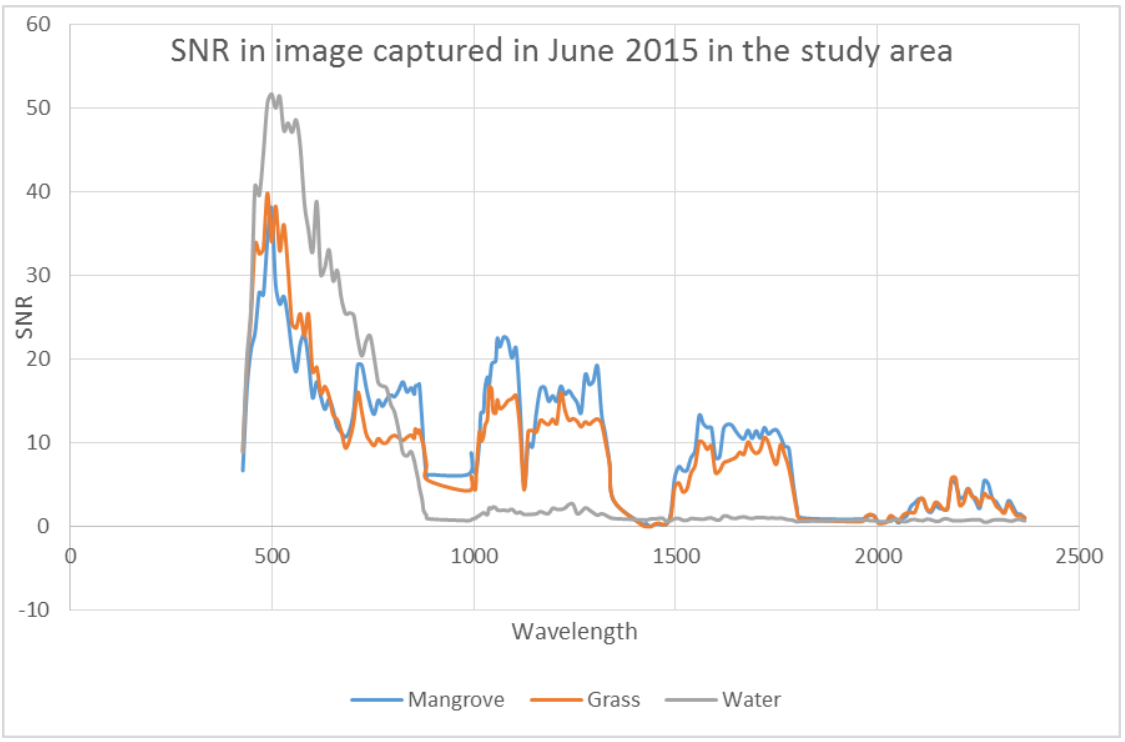

Figure 2-6: . SNR of EO-1 Hyperion data (winter season).

The SNR is higher in summer for all the three land covers (figure 2-5) and lower in winter (figure 2-6). The estimated SNR for both seasons is in good agreement with the predicted SNR for EO-1 Hyperion (Pearlman et al. 2003) . The SNR was one of the parameters that need to be estimated to establish the quality of images acquired by the sensors.

2.3.3 Principal Component Analysis

PCA was applied on the atmospheric corrected and a spectral subset of 56 bands. Depending on the amount of information and lack of gain of the variance in the increasing PCs, the initial intrinsic dimensionality is reduced to 16 components (figure 2-7).

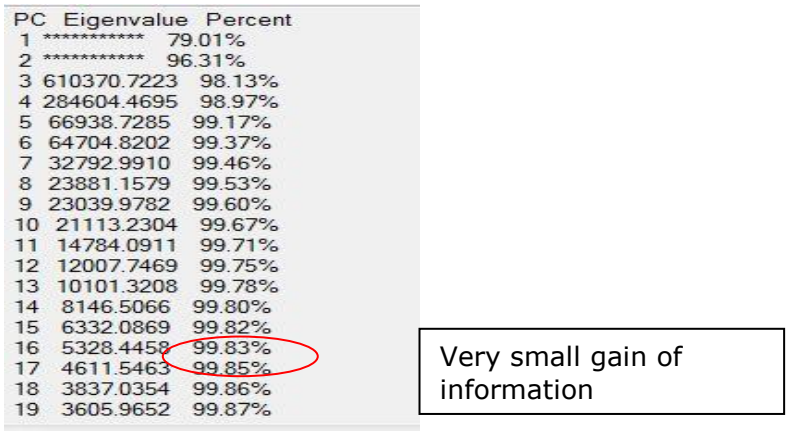

Figure 2-7: Percentage depiction of gain in variance with increase in PCs.

2.3.4 *Selection of Band for classification*

Based on the vegetation study and references (Thenkabail et al. 2004a; Thenkabail et al. 2004b), total of 16 bands (table 2-3) were selected for further classification.

Table 2-3: List of 16 selected bands for classification.

Worldview-2	EO-1 Hyperion			
Bands	Region of Spectrum	Selected Bands	Wavelength (nm) range	Importance as per (Thenkabail et al. 2004a; Thenkabail et al. 2004b)
Coastal	Visible	B9	436.99	Blue absorption peak, chlorophyll-a
Blue	510	B20 B23	548.92 579.45	Absorption pre-maxima, soil background
Green	580	B25	599.80	
Yellow	625	B26 B27	609.97 620.15	
Red	690	B29 B33	640.50 681.20	Absorption maxima, maximum chlorophyll absorption
Red Edge	745	B35 B39 B41	701.22 742.25 762.60	Sensitive to vegetative stress
NIR1	895	B42 B57	772.78 884.70	Correlation with Biomass
NIR2	1040	B79 B82 B86	996.63 1013.30 1033.50	Sensitive to moisture plant moisture stress
8 Bands		16 bands		

2.3.5 Analysis of radiance and reflectance spectra and supervised classification

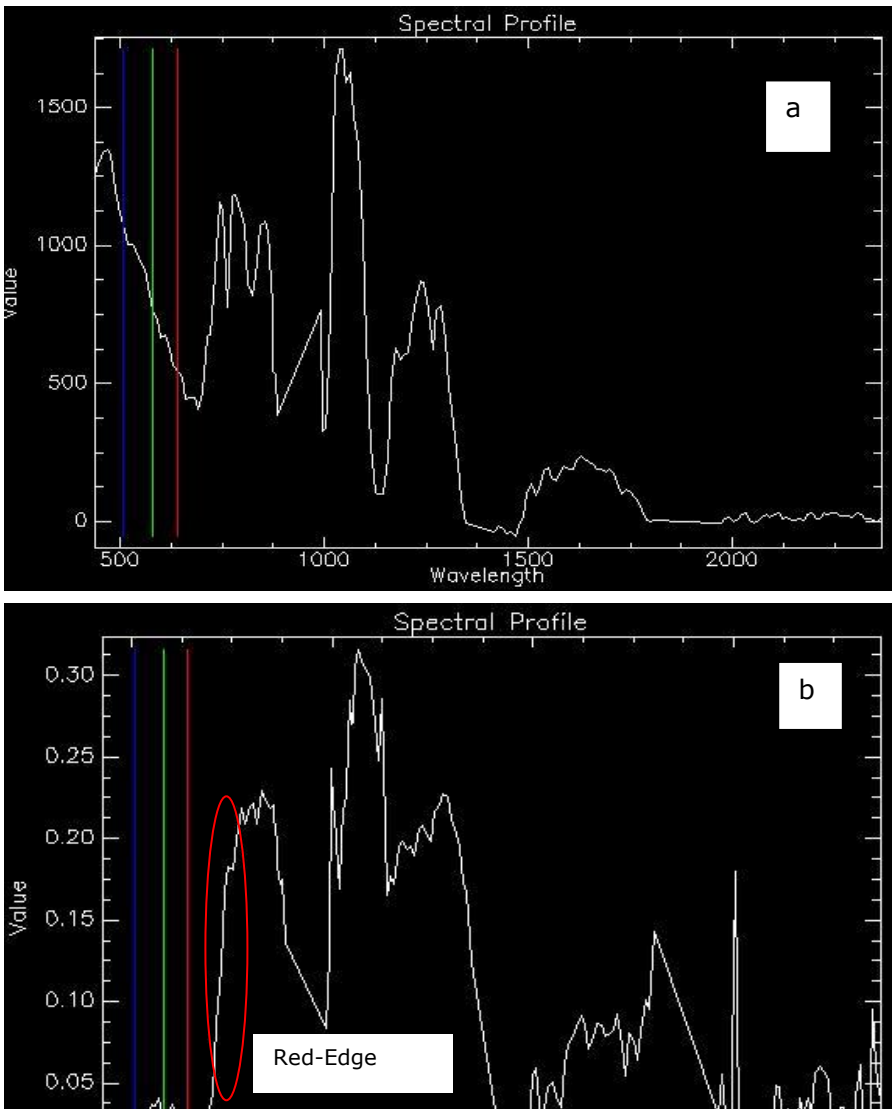

Figure 2-8 Radiance (a) and surface reflectance (b) spectra for healthy forest of EO-1 Hyperion data.

Radiance spectra of Hyperion data includes radiation reflected from the surface and affected by the source of radiation that is sun for optical imagery. From figure 2-7 (a), the

radiation spectra trend toward higher values at about 500 nm, because the spectrum of the sun peaks at about 500 nm and looks like the overall shape of the solar spectrum. That is for any quantitative analysis of multispectral or hyperspectral image data, radiance image is corrected to reflectance images. From surface reflectance spectra (figure 2-7 b), the spectra changed and red-edge part has smoothened sharply that is most important for vegetation spectral properties analysis

Table 2-4: Supervised classification of Hyperion data

Bands combination	OA_Training %	Kappa statistics	OA_Validation %	Kappa statistics
RGB and NIR1 (8band)	98.42	0.978	70.47	0.51
8 band and Coastal (10 Band)	98.42	0.978	70.43	0.51
10 Band and Yellow(12 bands) 12 The band and Red Edge (14 Band)	98.76 99.15	0.981 0.988	70.38 71.59	0.50 0.52
14 Band and NIR2 (16 Band)	99.27	0.988	71.66	0.52

Table 2-5: Mapping Accuracy (%) of Hyperion data for the validation classes in saltmarsh ecosystem.

Band	Forested wetlands	*Phragmites australis*	*Sporobolus virginicus*	Other marshes	Water
RGB and NIR1(8band)	77.23	66.27	55.44	62.57	67.37
8 band and Coastal (10 Band)	76.56	66.30	54.19	64.03	70.34
10 Band and Yellow(12 bands) 12 The band and Red Edge (14 bands)	77.46 78.86*	66.59 67.85*	55.44 57.30	64.38 64.57*	70.56 73.54*
16 Band together	79.12*	67.95*	57.34*	65.23*	73.87*

2.3.6 Analysis of radiance and reflectance spectra from Worldview-2 data

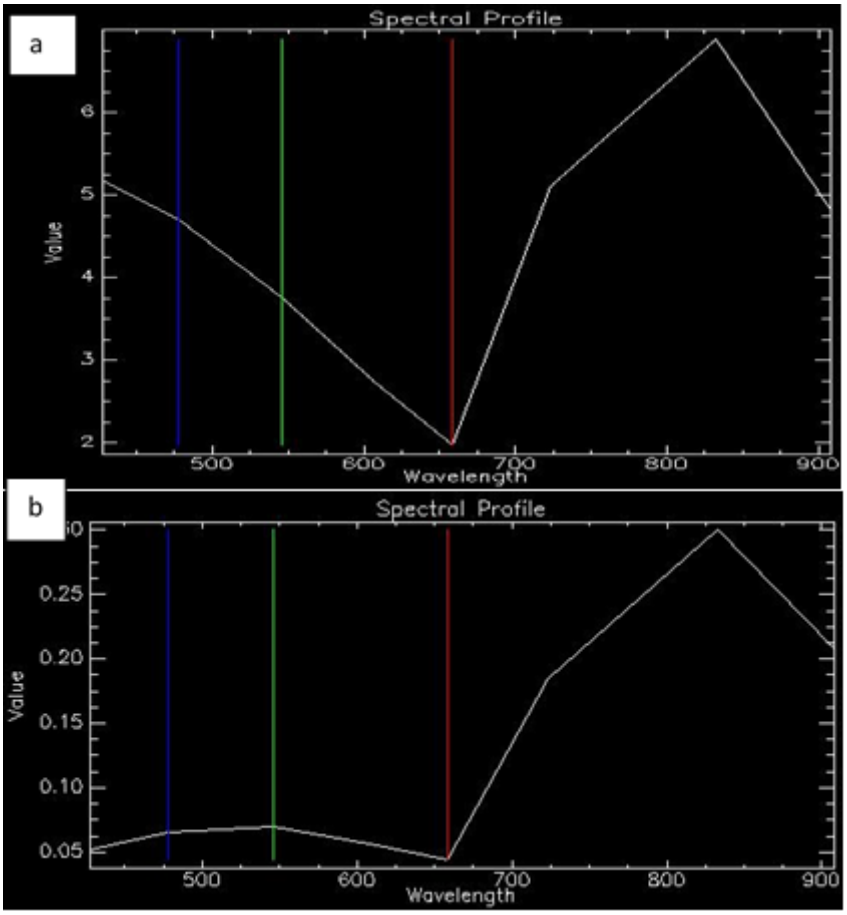

Figure 2-9: Radiance (a) and Surface reflectance (b) of healthy vegetation from healthy forest spectra.

At first radiance spectrum of different vegetation and land cover classes were visually observed to check their similarity and difference with the surface reflectance spectra. The radiance spectra (figure 2-9 a) shows high values within the blue and green part of visible wavelength due to aerosol scattering. But after FLASSH, the surface reflectance spectra (figure 2-9 b) is corrected and blue and green values are much lower and the chlorophyll

peak in the green wavelength is visible. Now, this spectrum is comparable with the corrected reflectance spectra of Hyperion data.

Table 2-6: Supervised classification of Worldview-2 data

The arrangement of bands for classification	OA_Training %	Kappa statistics	OA_Validation %	Kappa statistics
RGB and NIR1	97.61	0.969	72.57	0.56
RGB, NIR 1 and Coastal	97.58	0.964	72.03	0.56
RGB, NIR1 and Yellow	98.14	0.973	76.38	0.62
RGB, NIR1 and Red Edge	98.19	0.973	78.59	0.64
RGB, NIR 1 and NIR2	98.19	0.972	78.66	0.65
8 Band together	99.07	0.984	79.67	0.65

Table 2-7: Table 2.6 Mapping Accuracy (%) of Worldview-2 data based on the dominant classes

Band	Tree species	Phragmites	Sporobolus	Other marshes	Water
RGB and NIR1	81.23	68.43	74.68	66.57	78.86
RGB, NIR 1 and Coastal	82.56	69.60	78.58*	72.53*	79.02
RGB, NIR1 and Yellow	81.46	62.59	77.02	67.38	78.92
RGB, NIR1 and Red Edge	88.19	72.79	78.25	70.59	81.54
RGB, NIR 1 and NIR2	88.55*	73.45	77.58	71.66	84.09*
8 Band together	89.12*	78.74*	81.78*	72.67*	88.56*

It is clear from the training and validation dataset, with an increase in the number of the bands, the overall accuracy also increased except the coastal band. This might be due to the absence of seawater training data and location for validation. Overall accuracy for test site increased up to 7% with the combined 8 bands of Worldview-2 data.

2.4 Discussion

Not only was the Hunter region, SNR calculated from the other parts of Australia. Using the mean/standard deviation method, 4 Hyperion scenes from around the coastal region of Australia shows that there is a strong relationship between the acquisition time of year and the SNR of the Hyperion data. That was a good agreement with the findings of Kruse et al. (2003). Calculated SNR for Hyperion SWIR data is higher in the summer and lowest in the winter (Figure 2.5 and 2.6) that was also similar to the finding of (Kruse et al. 2003). Based on the SNR based data quality assessment, Hyperion L1_R data were radiometrically calibrated followed by FLASSH correction.

The variability at the green peak (~550nm), chlorophyll absorption (680nm) to approximate Red- Edge (720 nm) and water absorption features (945nm and 1190nm) arises due to the variation of pigment, water content and the structural components. These differences lead to different species or variation in health or phonological stage within the species. Compared to green grass (Green colour spectra) shown in Figure 2.4 b, the chlorophyll absorption of *Casuarina glauca* and *Phragmites australis* are weaker and they are structurally different. *Casuarina glauca* (Cyan colour spectra) is also different from *Phragmites australis* (red colour spectra). *Phragmites australis* can be identified at 2100 nm to 2300 nm (red spectra is distinct) due to cellulose content. The spectral shape, cellulose related features and the depth of the chlorophyll absorption (shallow depth compared to green and cyan) could be attributed to the tall less green *Phragmites australis*. The PPI algorithm didn't work properly to identify pure pixel for *Sporobolus virginicus* and *Sarcocornia quinqueflora* species. The main reason is *Sarcocornia quinqueflora land* cover fraction was very low compared to two saltmarsh species.

There are clear and visible differences between radiance and reflectance spectra of WV-2 and EO-1 Hyperion data that becomes visible after atmospheric correction. Øystein and Øivind (2012) proved that FLAASH corrected Worldview-2 image has a clearly lower blue component and an expected chlorophyll peak in the green band due to the correction of aerosol scattering. In a different study conducted by Yuan and Niu (2008) showed that Hyperion image showed the rich spectral information of objects after FLAASH correction.

Based on the PCA of FLAASH corrected reflectance data, only 16 PCs were found that contained most of the information. It was similar to the findings of Chauhan et al. (2011) and Pervez and Khan (2015). From 155 atmospherically corrected band, Pervez and Khan (2015) showed that only first 10 PCs contain more than 99 % of the information. Chauhan et al. (2011) segregated first 13 bands that contain 97 % information from 168 bands of Hyperion data. In our study, among the 56 bands of atmospherically corrected Hyperion data, 79.01% variability was contained by the first PC, 96.31% variability was contained by PC2 and likewise, 16 PCs contains up to 99.83% variability due to the application of PCA. PCA highlights the redundancy in data due to similar responses in some wavelength and reduces the dimensionality of data by decorrelation. When 16 selected bands of

Hyperion data were compared with the 8 bands of Worldview-2 for saltmarsh classification, the overall classification accuracy has increased in both cases after adding band orderly. But the overall accuracy obtained from Worldview-2 was higher than that from the EO-1 Hyperion image. Table 2.5 shows that OA % for Worldview-2 was increased from 72 to 79 while for Hyperion it increased from 70.47 to 71.66. Considering the significance test with z values and kappa statistics at 95% confidence interval, Worldview-2, classification accuracy was higher than Hyperion data. These findings differ from the findings of Kumar and Sinha (2014). This is might be due to the spectral properties of high spatial resolution data and ground scene of the study area. Kumar and Sinha (2014) used Quick bird images that have 4 multispectral bands with 2.4 m resolution. Whereas Worldview-2 images that used in our study have 1.84 m spatial resolution with 8 bands in the multispectral mood. Moreover, they used all of the bands of Hyperion data in their study. But in our study, we used only 56 Hyperion band (49 VNIR and 7 SWIR) to compare with the wavelength of Worldview-2 images. We also segregated a number of bands to reduce the redundancy of information. Finally, 16 bands based on different literature were used for classification purposes.

2.5 Conclusion

Based on pre-processing, the data reduction algorithm, PPI and other investigation we conclude that EO-1 Hyperion hyperspectral data shows potential for the identification of the endmember from wetland components. However, it was found that two important factors impede the PPI endmember extraction procedure and subsequent supervised classification of hyperspectral data. These are the lack of pure pixels and the small patch size on the ground. PPI endmember extraction algorithms assume the presence of pure pixels and its threshold level based training pixels and image pixels. For spaceborne hyperspectral data with a typical 30-m spatial resolution, it is very difficult to find pure pixels in a wetland ecosystem. A high spatial resolution data could resolve that issue, specifically, when filed patch size of saltmarsh is not large enough to match with 30m pixel size of image data. This research describes the importance of SNR for data quality assessment and PCA for data reduction for EO-1 Hyperion data. Based on the VNIR of multispectral broadband and hyperspectral narrowband data this research explores the potentiality of spatial resolution over spectral resolution. Classification accuracy improved significantly in both cases after adding bands orderly. But overall accuracy was higher in case of Worldview-2 due to high spatial resolution and small patch size of species on real earth condition. The result of the current study can be applied to any future research relates to VNIR for the improvement of classification accuracy. Although this research ignores the importance of SWIR, next chapter will focus on the application of Landsat 8 OLI for spectral profile analysis for saltmarsh species that has two SWIR bands. Due to the small patch size and difficulties in access to the study area, it was really difficult to collect enough sample points for calibration. So the region of interest based training sampling can be a recommendation for further research. In addition, spectral separability analysis is recommended to test the separability between species. As data collection is a challenging task, advance machine learning classifiers can be considered in future work as they work based on small sample size and multimodal distribution data.

Chapter 3
Spectral Separability and spatial sampling unit analysis of saltmarsh species class using Worldview-2 and Landsat 8 OLI data

This chapter is based on:

Comparative analysis of Worldview-2 and Landsat 8 for coastal saltmarsh mapping accuracy assessment. Proc. SPIE 9864, Sensing for Agriculture and Food Quality and Safety VIII, 986409 (May 26, 2016); doi:10.1117/12.2222960.

Abstract

The distribution of classes on feature space is an important parameter that often has overlaps which reduce the accuracy of classification. To solve this problem, the concept of spectral separability is introduced. So the class separability varied with sampling design and the spatial unit (region of interest or pixel). Spectral separability was tested based on divergence and Jeffries-Matusita (JM) distance to measure class separability. The spectral profile was analysed based on two different sampling criteria followed by a supervised classification using Landsat 8 OLI and Worldview-2 data. Two different sensors, High Spatial Resolution Worldview 2 data and Coarse Spatial resolution, Landsat 8 were selected for this study. Among the selected vegetation types some patch was fragmented and close to the spatial resolution of Worldview 2 data and some patch was more than 30-meter resolution of Landsat 8 data. The main objective was to check the spectral profile originated from the region of interest and pixel-based sampling. Based on the spectral profile, Mangrove and She-oak were clearly separable using Landsat 8 data. But from Worldview-2 data, it was not possible to separate because error bars were overlapped. A similar trend was found for two saltmarsh species with Worldview-2 data. However, when pixel based sampling was considered, red band information was effective to differentiate the spectral profile of Phragmites australis and Sporobolus virginicus. Based on the region of interest based sampling (ROI), supervised classification provided 88.72% overall accuracy that was higher than the OA of Landsat 8 OLI. However, classification error was evident due to the region of interest based sampling that includes heterogeneous pixel within the homogenous category. Due to the scattered distribution of saltmarsh species, pure homogenous pixels for training data were very limited. Therefore, advanced machine learning algorithm has been recommended for future research work.

Keywords: Saltmarsh, Worldview 2, Landsat 8 OLI, spectral separability, a region of interest (ROI).

3.1 Introduction

Saltmarsh usually lives with dominant mangroves in the wetland ecosystem. This causes a problem in selection of scale and radiometric properties of the remotely sensed data when performing classifications. The scale or spatial resolution is the projected area on the ground associated with the radiance measurement of the sensors. It is the smallest distinguishable spatial unit (ground resolution element) recorded in a remotely sensed data. For example, the spatial resolution is 2 m for Worldview-2 data and 30m for Landsat 8 OLI or EO-1 Hyperion data. Sometimes the spatial resolution denotes the ground sampling distance in an image after image re-sampling that can be varied from the actual spatial resolution recorded by the sensor. Spectral resolution means the range over the electromagnetic spectrum the energy is measured and recorded by the sensor. The use of higher spectral resolution images usually improves the capability of detecting spectral variability within a land cover by increasing the discrimination capacity (Almeida and Filho 2004). But a spatial resolution may be more important than spectral resolution when spatial extent of land cover types is an issue. However, both spatial and spectral resolution are two fundamental characteristics of a remotely sensed image for any application, similar to classification or feature extraction from feature space.

The distribution of classes on feature space is an important parameter that is related to the classification accuracy. This is because class probabilities are produced based on the distribution of classes in feature space. However, this distribution often has overlaps which reduce the accuracy of classification. To solve this problem, the concept of spectral separability is introduced. This concept indicates how well two classes are separated. This separability concept is classical in pattern recognition and independent of the coordinate system(Fukunaga 2013; FUKUNAGA 1990) and has an effect on classification accuracy.

Remote sensing image classification accuracy is affected by other two factors: firstly, the influence of boundary pixels and secondly, a finer spatial resolution that increases the spectral –radiometric variation of land cover types (Markham and Townshend 1981). Optimum spectral and spatial resolution determination for vegetation mapping has been an ongoing area of research in remote sensing (Curran and Atkinson 1999). It becomes more challenging when dominant and fragmented plant species patches are distributed together in the same community, similarity to our study site.This study site supports an extensive temperate saline coastal wetland system comprising mangrove, saltmarsh and mixed mangrove–saltmarsh habitats (Rogers et al. 2014). About 4,257 ha of these wetlands are formally protected within the Hunter Wetlands National Park and this includes over 2,926 ha listed as internationally important under the Ramsar Convention in 1984. However to facilitate agriculture and industrial production, a network of 176 levees, culverts and floodgates that were constructed in the period 1950–1980 under the Hunter Valley Flood Mitigation Scheme (Winning and Saintilan 2009). As result wetland areas

behind levee banks have transitioned from saltmarsh to brackish reed swamps dominated by *Sporobolus virginicus, Schoeloplectus subulatus* and *Phragmites australis* (Winning and Saintilan 2009). Due to this, the whole wetland and the surrounding part is a place of scientific research based on different objectives. The main interest of this stud was to know the current status of theses degraded saltmarshes with respect of the other co-habitants, for example, mangrove and She-oak and to find out optimum mapping accuracy for each saltmarsh species from remote sensing data. Species classification based on spectral properties of different vegetation classes within the wetland environment is a challenging task. Saltmarsh vegetation types may possess similar spectral signature in the remotely sensed image, and spectral resolution of data may be insufficient to detect saltmarsh class if spatial resolution is coarse. In the previous chapter, it was found that found that high spectral resolution data suffer from the redundancy of information, and all narrow bands are not effective to give enough information. However, due to the focus on the other objectives, spectral separability analysis was not tested in the previous chapter. In addition, the effect of the spatial unit, an i.e. region of interest (ROI) or single pixel issues were not considered in the previous chapter as it was not related to the hyperspectral data. Therefore, in this study, commonly used broadband multispectral data Landsat is introduced. Then the performance of High-Resolution Satellite Imagery (HRSI) Worldview-2 and recently launched coarse spatial resolution Landsat 8 OLI were compared. Both sensors have some similarities in their spectral coverage, however, their spatial resolution (pixel) is very different. This paper aims to examine the effects of the spatial unit (ROI or pixel) based sampling on spectral separability analysis followed by thematic map accuracy based on spatial resolution irrespective of spectral resolution. This study also explores the importance of moderate resolution (30-meter) broadband data for two dominant species, Grey Mangrove (*Avicennia marina*) and She-Oak (*Casuarina glauca*).

3.2 Dataset and methods

3.2.1 Input Data

Two different platform satellite imagery of Worldview-2 and Landsat 8 OLI were used for this research. Worldview-2 has a pixel resolution of 0.46m in the panchromatic mode and 1.84 m resolution in the multispectral mode whereas recently launched Landsat 8 OLI consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. New band 1 (ultra-blue) is useful for coastal and aerosol studies. Another new band 9 is useful for cirrus cloud detection. The resolution of Band 8 (panchromatic) is 15 meters. Thermal bands 10 and 11 are useful in providing more accurate surface temperatures and are collected at 100 meters. Approximate scene size is 170 km north-south by 183 km east-west (106 mi by 114 mi). In our study, we did not use thermal and cirrus bands (table 3-1).

Table 3-1: Specification of Worldview-2 and Landsat 8 OLI data.

Worldview-2			Landsat OLI imagery		
Bands	Central Wavelength (nm)	Wavelength range (nm)	Bands	Central Wavelength (nm)	Wavelength (nm) range
Coastal	427	400-450	Coastal	443	433 - 453
Blue	478	450-510	Blue	482	450 - 515
Green	546	510- 580	Green	562	525 - 600
Yellow	608	585-625	Red	655	630 – 680
Red	659	630-690	NIR	865	845 -885
Red Edge	724	705-745	SWIR 1	1610	1560 -1660
NIR1	833	770- 895	SWIR 2	2200	2100 - 2300
NIR2	949	860-1040	Pan	600	500 - 680
Pan	627	447- 808	CIRRUS	1370	1360-1390

3.2.2 Image acquisition and Image processing

For Worldview-2 imagery, the raw Digital Number (DN) has been converted at sensor radiance using Worldview-2 calibration utility, available in ENVI V4.6 and greater. The image was acquired in May 2015 that is the transition of the season from summer to winter in NSW, Australia. Based on the initial analysis of the DN value of the image we found that coastal band (lower spectral region) is affected by atmospheric scattering. To compare the effect of atmospheric correction of image spectrum we applied atmospheric correction method, The Fast Line-of-sight Atmospheric Analysis of Spectral Hypercube (FLAASH). Afterwards, the WorldView-2 imagery was registered to Map Grid of Australia (MGA94) Zone 56 using a high-resolution aerial photo of the local council, Land and Property information, NSW, Australia (https://maps.six.nsw.gov.au/). A first-order polynomial transformation was applied and RMSE values estimated for this transformation was 0.32 pixels. The first order polynomial was selected because this wetland area has a generally flat terrain. However, the RMSE was considered well enough considering the nominal GPS measurement error of 3-4 m that we obtained in the field.

Similarly, the DNs of the Landsat OLI data were first converted to at-sensor radiance by using the radiometric calibration parameters. The FLAASH algorithm was then used to convert radiance to reflectance. Same aerial photo (SIX Map) was used for rectification of Landsat OLI imagery. A first-order polynomial transformation with the nearest neighbourhood resampling was applied and RMSE values estimated for image transformation were about 0.87 pixels.

3.2.3 Processing of field data

There are three different strata (Tree, saltmarsh and water) at the study site. Randomly collected ground reference sample from different strata was divided into two groups, the calibration and validation set. This is the basic criteria to remove any possible bias that could be caused by using the same set of pixels to calibrate and to validate the classifiers. It is known that Landsat OLI data were hampered by the low spectral dimensionality and spectral resolution (Xiao and Moody 2005). In this circumstances, to deal with medium resolution (30 m) Landsat OLI data two dataset was prepared for classification purposes. One dataset is prepared based on the overlaying of field sample vector file (stratified random sampling) over Landsat OLI and prepare region of interest (a group of pixels/polygon) for each class based on homogeneity. The spatial distribution of those sample points are shown is chapter 1 (figure 1-2). Those reference points were used to identify the homogenous pixels for each species. The homogeneity of these pixels cluster can be tested with the Principal Component Analysis (PCA). Because PCA used the highly uncorrelated bands in a 2-dimensional scatterplot (figure 2) and could be used to identify the individual endmembers of multiple surface components (Johnston and Barson 1993; Smith et al. 1985). According to Johnston and Barson (1993) and Smith et al. (1985), in a mixture of three substances, the first two components produced a triangle (Figure 3-1) in a scatterplot to locate the endmembers at the corner of the triangle. This was the best approach to make our training and test data unbiased as much as possible. For Worldview-2 data, total 2228 training and 1625 test data were used for eight different classes. On the other hand ROI, 638 pixels were used based on principal component analysis to calibrate the model for Landsat 8 OLI and 529 pixels were used to validate the result. Another dataset was used to test the individual pixel separability analysis that is developed from the fieldwork and individual pixel for each species.

3.2.4 Class separability

There are several methods to identify class separability. For example, Jia and Richards (1999) mentioned the divergence and Jeffries-Matusita (JM) distance to measure class separability. In this research Jeffries-Matusita (J-M) distance was used to measure the separability between the 8 classes chosen. The value of this distances varies from 0 to 2.0 where 0 indicates classes are same and 2.00 indicates they are very well separable.

3.2.5 Selected classifier

Maximum Likelihood Classifier (MLC): MLC is one of the statistical classifiers. It is based on that depends on the idea that the distribution of the data in each class is normally distributed and called as a parametric classifier. This classifier assumes that the statistics for each class in each band are normally distributed and calculates the probability that a given pixel belongs to a specific class. Unless a probability threshold is selected, all pixels are classified based on the probability of normally distributed data. Each pixel is assigned

to the class that has the highest probability. If the highest probability is smaller than a threshold, the pixel remains unclassified. More details of MLC can be found in Otukei and Blaschke (2010).

3.2.6 Accuracy Assessment and kappa analysis

Accuracy assessment results were discussed based on the confusion error matrix, overall map accuracy and kappa values. For individual class, producer and user accuracy were computed based on the dominant class in each reference plot (Richards 1996; Stehman 1997; Story and Congalton 1986)

3.3 Results and discussion

3.3.1 Principal Component Analysis

The PCA eigenvalue showed that the first three PCA components accounted for almost 99.50% of the total variance for Worldview-2 data and 99.76% for of the total variance for Landsat 8 OLI data. However, if we consider the pixel resolution of Worldview-2 data then rest 0.50% variance is really important for 2X2 m pixels. But we can ignore rest 0.25% variance (table 3-2) based on 30x30 m pixels resolution for Landsat OLI data to select independent endmember dataset for classification. In this perspective, our calibration data from Landsat OLI extracted a maximum number of endmember (for Landsat 8 it is 7) from uncorrelated PCA bands. Existing studies suggest that the pixels located on the corner of the scatterplot (figure 3-1) can be treated as endmember (Qu et al. 2014).

Table 3-2: Percentage depiction of gain in variance with an increase in PCs in both dataset

	Worldview-2			**Landsat OLI**		
PCs	**Eigen value**	**% Variation**	**Cumulative percentage**	**Eigen value**	**% variation**	**Cumulative Percentage**
1	0.063954	96.9822	96.982288	0.023215	93.9802445	93.9802445
2	0.001278	1.9380	98.920296	0.001256	5.08460853	99.0648530
3	**0.00039**	**0.5944**	**99.514739**	**0.000174**	**0.7043964**	**99.769249**
4	0.000195	0.2957	99.810445	0.000029	0.11739940	99.8866488
5	0.00005	0.0758	99.886267	0.00002	0.08096510	99.9676139
6	0.000032	0.0485	99.934793	0.000007	0.02833778	99.9959517
7	0.000024	0.0363	99.971187	0.000001	0.0040482	100
8	0.000019	0.0288	100			

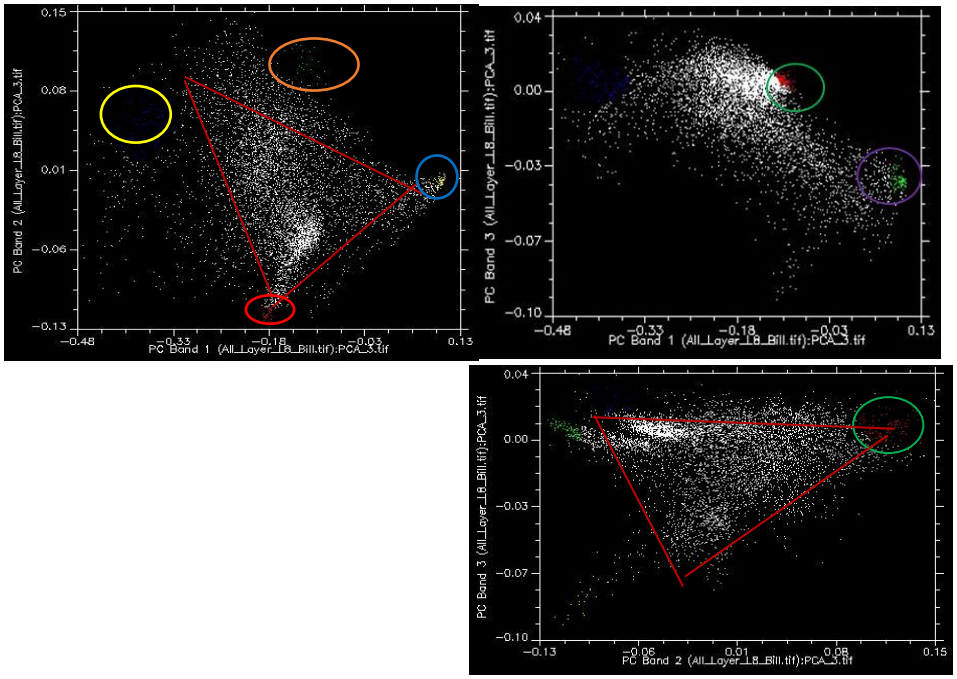

Figure 3-1: Principal Components Analysis (PCA) based endmember selection for classification of Landsat 8 OLI data. Scatterplot of the three PCs from Landsat OLI data to select seven endmember: (Clockwise) PCA 1vs PCA2: Mangrove, Grass, Phragmites australis, Water; PCA 1 vs PCA3: Casuarina glauca, Sporobolus virginicus, Water; PCA 2 vs PCA3: Phragmites australis.

3.3.2 Spectral profile Analysis

Based on the spectral curve originated from the region of interest (ROI / polygon), it is very clear that all saltmarsh species are overlapped with each other within all bands of the spectrum. But Mangrove and She-oak (*Casuarina glauca*) tree species are clearly separable from each other up to bands 5 (figure 3-2). Because Mangrove and She-oak are two dominant tree species cover a large extent of the study area that is easily detectable from 30-meter pixels of Landsat data. Similarly, band 1-4 are a clear indication to separate Marshy wetland from the water. On the other hand for Worldview-2 data (figure 3-3), reflectance properties of water in bands 6-8 (Red Edge, NIR1 and NIR 2) and reflectance of marshy wetland in band 7 are significantly different from other classes. The perennial grass is unique in band 3(Green) and 6 (Red Edge). However, in most of the bands, the error bars of Mangrove, She-Oak, *Phragmites australis*, and *Sporobolus*

virginicus overlapped with each other. Among these species, the spectral separation is not very clear as there is overlap in the reflectance region.

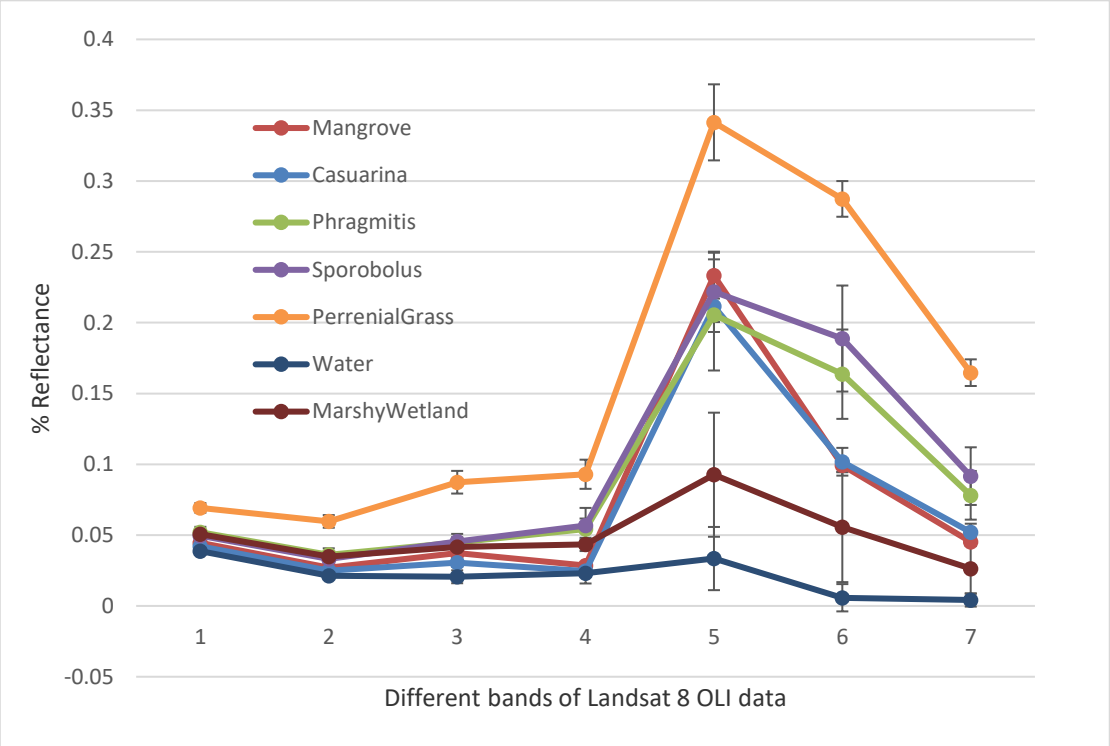

Figure 3-2: Different classes extracted from 7 bands of Landsat 8 OLI data

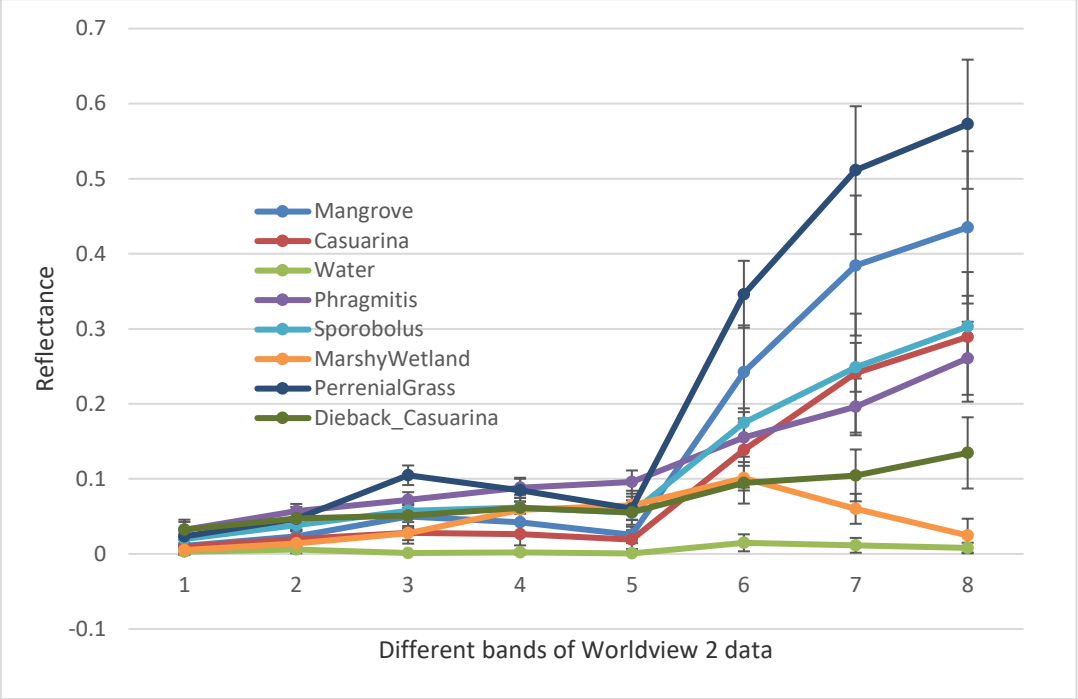

Figure 3-3: Different classes extracted from Worldview-2 data

The box-whisker plot developed from pixel-based sampling from Worldview-2 indicates the spectral variability and data distribution pattern among and within 8 classes (figure 3-4). Only annual grass (GR) is clearly separable in green, red edge and NIR bands. Reflectance properties of water (WA) in the last 3 bands (Red Edge, NIR1 and NIR 2) and reflectance of wetland soil (WS) in band 8 are significantly different from other classes. Within two tree species (Casuarina=CA and Mangrove=MA) the separation is not so clear, there are considerable spectral overlaps in different bands. Two saltmarsh species (*Phragmites*= PH, *Sporobolus* = SP) are clearly separable in the red band. The saltmarsh species also show band-specific within-species variance. For example, the variance of *Phragmites* (PH) in the coastal band is quite large and in NIR2 is relatively small. Based on the overall box-whisker plot, it is clear that there are no outliers for any species in any band. Moreover, the median is the centre of each box indicating that data are normally distributed.

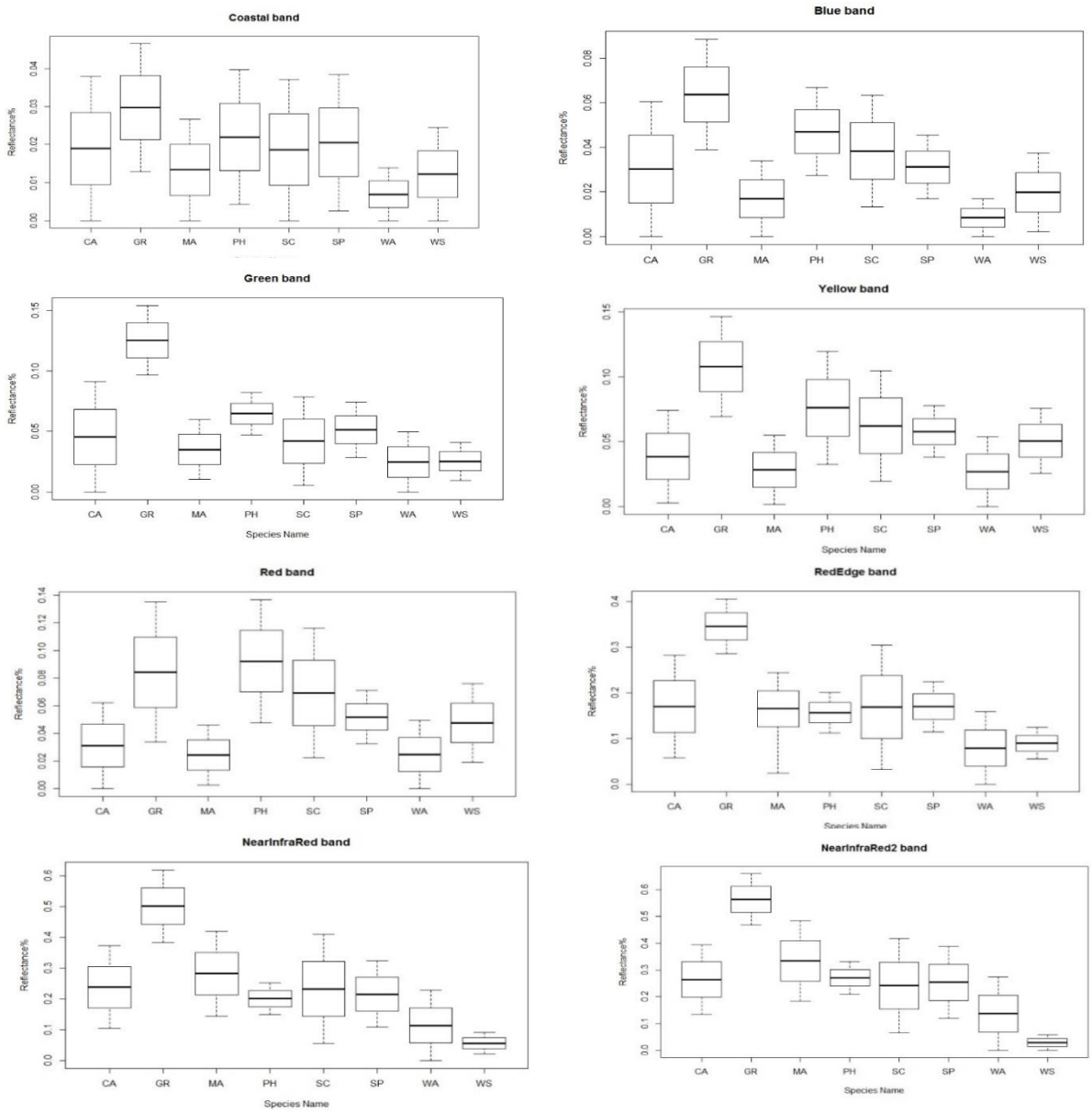

Figure 3-4: Box-whisker plots of median reflectance values of the 8 Worldview 2 bands for different wetland species derived from the reflectance pixels of the study area

3.3.3 Classification Result for Landsat 8 OLI

The confusion matrix (table 3-3) of the MLC classifier originated from all the bands of Landsat 8 OLI data indicates that the producer and user accuracy for Mangrove, Casuarina and Grass species are higher than others species. Although overall classification accuracy is lower for all classifier when it was compared with that of Worldview-2 data (table 3-4). But the main findings is that the two dominant species of Mangrove and She-Oak are well classified from the Landsat 8 image. However, it was not possible to collect enough sample for dieback disease using coarse resolution data and has been removed from Landsat 8 OLI.

Table 3-3: Confusion matrix of MLC classification of Landsat 8 OLI data

MLC										
Class	A	B	C	D	E	F	G	Total	EC	UA
A	**54**	1	0	0	0	0	7	56	1.78	98.21
B	15	**65**	0	9	0	0	2	91	9.89	90.10
C	0	0	**52**	11	0	0	0	70	15.71	84.28
D	0	0	14	**39**	0	0	5	58	22.62	77.37
E	1	0	0	0	**81**	5	0	87	5	95
F	3	1	0	0	4	**79**	1	87	6.15	93.85
G	0	0	8	7	0	2	**63**	80	10	90
Total	73	67	74	66	85	86	78	**529**		
EO	20.54	20.37	23.52	21.60	4.70	5.81	7.69	OA=82. 04% Kappa =0.79		
PA	79.45	79.62	76.47	78.39	95.29	94.18	92.30			

Class key: A= Water, B = Marshy Wetland, C = *Phragmites australis*, D = *Sporobolus virginicus*, E= Mangrove, F = Casuarina, G = Perennial Grass, EO = Error of Omission, EC= Error of Commission, UA= User Accuracy, PA= Producer Accuracy, OA = Overall Accuracy

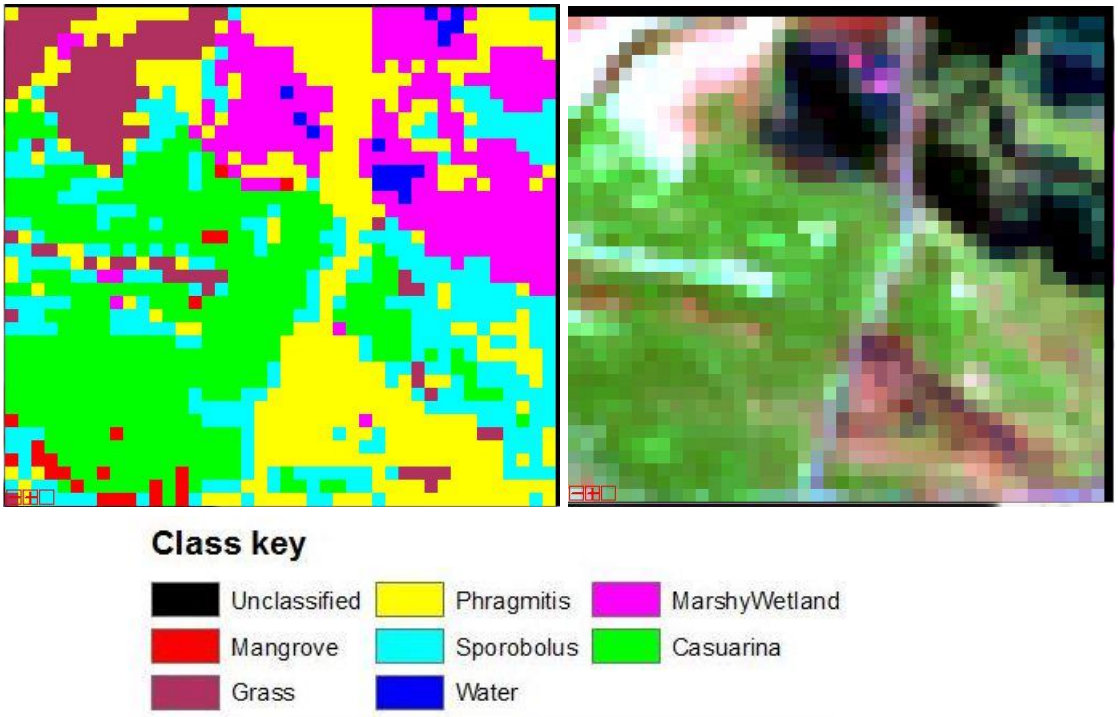

Figure 3-5: (right) True color image of the part of a study area and (left) MLC classificationof Landsat 8 OLI.

From the figure 3-5, it is clear that dominant species *Casuarina glauca* (She-oak) has been identified with a high accuracy as it has a larger area that can be mapped by 30x30m resolution data. On the other hand, small patch size of *Phragmites australis* has been misclassified that is also reflected when mapping accuracy was calculated. This misclassification raised due to the spatial resolution of Landsat 8 OLI data. Although confusion matrices derived from Landsat 8 is showing higher accuracy for different small fragmented saltmarshes, this accuracy is not always acceptable. Because a 30-meter pixel is a combination of different species/land cover. These mixed pixels have not properly classified by this broadband data. However, ROI based sampling effectively classified dominant tree species (Mangrove or She-Oak) that is very clear in figure 3-5.

3.3.4 Classification Result for Worldview-2 data

The confusion matrix (table 3-4) of the MLC classifier indicates that there is a high incidence of misclassification of Mangrove and She Oak from Worldview-2 data. This misclassification mostly comes from the similarity of the spectral properties of this tree species.

Table 3-4: Confusion matrix of MLC classification of Worldview-2 data

Class	A	B	C	D	E	F	G	H	Total	EC	UA
A	**248**	56	0	0	0	0	4	0	308	57.2	42.7
B	58	**128**	0	0	0	0	4	0	190	60.1	39.8
C	1	0	**366**	0	0	1	0	0	368	1.27	98.7
D	0	0	0	**286**	5	0	7	1	299	12.3	87.7
E	0	0	0	10	**129**	0	6	4	149	18.6	81.3
F	1	3	3	2	1	**130**	2	0	142	8.45	91.5
G	0	0	0	7	5	0	**30**	0	42	18.9	81.0
H	0	0	0	1	1	0	0	**125**	127	1.57	98.4
Total	308	187	369	306	141	131	53	130	**1625**		
EO	18.8	31.5	0.8	37.2	10.7	1.85	43.4	3.1	OA=88.73% Kappa =0.82		
PA	81.1	68.4	99.1	62.7	89.2	98.1	56.6	96.9			

Class key: A= Mangrove, B = Casuarina glauca, C= Water, D = *Phragmites australis*, E = *Sporobolus virginicus*, F= Marshy wetland, G = Dieback Casuarina, H = Perennial Grass, EO = Error of Omission, EC= Error of Commission, UA= User Accuracy, PA= Producer Accuracy, OA = Overall Accuracy

Table 3-5: Accuracy assessment and kappa statistics

MLC	Worldview 2	Landsat 8 OLI
Overall accuracy	88.73	82.04
Kappa coefficient	0.82	0.79
Variance	0.000048	0.00027

Table 3-6: Mapping accuracy assessment based on each class

Class	Worldview-2	Landsat 8 OLI
Mangrove	**66.48**	**89.01**
Casuarina	**67.36**	**84.04**
Water	97.34	74.32
Marshy Wetland	98.48	69.89
Phragmites	91.37	56.62
Sporobolus	84.86	45.88
Dieback Casuarina	42.10	
Grass	99.21	66.31

With Worldview-2 data, MLC produces lower MA for classes those have overlapping information in the electromagnetic spectrum, like as Mangrove and She-Oak. A similar trend was observed for *Phragmites australis* and *Sporobolus virginicus*. But MLC produces higher mapping accuracy for classes those have a fairly homogeneous composition of the

target classes at the 2X2 m spatial resolution. Because homogenous distribution results in relatively normal distribution of the reflectance values of that class in each of the spectral bands. For instance, water and grass were clearly separable and based on probabilistic parametric classifier MLC was giving the highest MA for these two classes. Similar studies in wetland areas have also delineated water body from other wetland features (Davranche et al. 2010).

On the other hand, the Overall Accuracy derived from Landsat 8 OLI was 82.04%. Producer Accuracy and User Accuracy were higher for Mangrove, She-Oak and grass due to their homogenous patch in the study site. These characteristics (higher spatial extent) played an important rule to improve Producer Accuracy and User Accuracy for these three classes. The main reason is that with coarse spatial resolution, dominant species with high spatial extent become easier to be defined. When we calculated the mean and standard deviation from different homogenous pixels for these two species we found a very low standard deviation. Due to this the error bar was not overlap within the first 5 bands for these two species. Contrary Worldview-2 (2m spatial resolution) data try to find out more chemical variation within visible to NIR due to more number of bands within the same spectrum. To some extent, within 2 m spatial resolution their leaf pigment or vigour might be similar and creates a confusion to classify it at this scale. But when it calculated the mean value within 30 m scales, it was clearly separable by Landsat OLI data. For this reason, the MA was lower for these two tree species. These findings also supported by Roth et al. (2015) as they got higher OA for some species when spatial resolution increased from 12 m to 20m. Our findings from Landsat 8 OLI are also comparable with (Chen et al. 2004) when they found a positive correlation between NIR and green bands with the spatial resolution of the image. They found that when the resolution becomes coarser, the correlation coefficient between the two bands is supposed to increase. In a similar way, Herold et al. (2004) compared simulated coarse resolution broadband IKONOS and AVIRIS data for classifying land cover and found that IKONOS was more appropriate than AVIIRS.

Individual pixel based separability index (Appendix-I) based on Jeffries-Matisuta (J-M) and Transformed-Divergence (T-D), indicate that *Phragmites australis* and *Sporobolus virginicus* are clearly separable (1.99). Similarly perennial grass, water and wetland are also separable from each other as their values are 1.99 to 2.00. In most of the cases, Mangrove and Sho-oak are not clearly separable as the value ranges from 1.84 to 1.98. It means in some cases they both are separable but in some cases not. However, it is very clear that individual pixel based sample collected from small patch size from the homogenous area (2- meter of Worldview-2) help to improve saltmarsh classes. Therefore, further research may be focused on the application of advanced algorithms based on pixel-based sampling originated from the small patch size of our study area.

3.4 Conclusion and recommendation

This study examined the distribution of fragmented saltmarsh in the context of two other dominant species, Mangrove and Casuarina. The effect of various spatial resolution on the accuracy of classifying the dominant vegetation species was investigated. The major findings are summarized as follows:

1. Highs spatial resolution Worldview-2 data with 8 spectral bands are used to make a map for fragmented saltmarsh where the minimum patch size is at least 2m by 2m for each species.
2. Coarse resolution Landsat 8 OLI data work well for classifying the dominant species when they spread over a wide extent. This is a good indication to map mono-species mangrove at community or regional level with Landsat 8 OLI data.
3. The region of Interest (ROI) based sampling is helpful to map a dominant species from coarse resolution data but ROI or polygon based sampling raised classification error for small fragmented species, i.e. *Phragmites australis*.

Laboratory analysis can be carried for leaf pigments, Chlorophyll and other nutrients that are related to plant reflectance. So that maximum information can be extracted from more bands of Worldview-2 data to remove overlapping information between two species. Pixel-based classification with the advanced machine learning classifiers can provide higher accuracy. So further research may be carried out based on the advanced machine learning algorithms.

Chapter 4

Scope for saltmarsh classification and biomass estimation using Support Vector Machine (SVM) and Random Forest (RF): A review

This chapter is based on

The article that has been submitted to the Special issue on "Marine Protected Areas: Science, Policy & Management" of the International Journal of Estuarine, Coastal and Shelf Science. This is a review article.

Abstract

Maximum likelihood (MLC) is the most common supervised classification method used in classification that follows a normal distribution. However, in most of the cases, the normal distribution is the main violation of RS data because the nature and causes of spatial variation in images are not always easily understandable that makes the application of MLC more difficult and inaccurate. In addition, adequate ground truth information and collection of enough sample points for training and validation are also impractical for wetland ecosystem that are the requirements for MLC. Therefore, the application of machine learning algorithms is increasingly popular for land cover mapping as they can handle small sample size and multi-modal distribution of data. Two machine learning algorithms, Support Vector Machine (SVM) and Random Forest (RF) have been reviewed in respect of land cover mapping and a special attention to wetland ecosystem. All types of remote sensing data including multispectral, hyperspectral, microwave RADAR and LiDAR data have been reviewed within the two broad categories of Machine learning classifiers. Although SVM related remote sensing works have been reviewed by Mountrakis et al. (2011). New articles have been published since then, and new methods have been used for SVM. Very recently, Belgiu and Drăguţ (2016) reviewed Random Forest and its application in remote sensing. But the potentiality of ensemble classifiers for a specific ecosystem like saltmarsh and wetland ecosystem need to be reviewed. That is why, emergence and modification of machine learning algorithms and techniques in recent years necessitate such a review, which will be highly valuable for guiding or selecting a suitable classification procedure for a specific ecosystem. In respect of the sensitivity of SVM application for classification and biomass modelling, t has been reviewed from the literature that the ability of 'small training size' and handle 'multimodal distribution' for model calibration are two important features of SVM. In addition, SVM does not encounter any over-fitting problem. As the collection of reference data for saltmarsh wetland classification is a challenging task, SVM might be a suitable alternative that can deal with a small sample size and can minimize the overfitting problem of biomass prediction model. However, kernel selection and other parameters (epsilon, cost and gamma) optimizations are major limitations that require time, skills and high computational cost. Random forest (RF) has some excellent advantages that make it popular in the field of classification and regression modelling. High classification accuracy, a novel method of determining variable importance (feature selection), fewer parameters (only two) are advantages of RF. Random Forest (RF), has reduced the problem of the multicollinearity issue that is a problem of multiple regression model, commonly used algorithm for biomass estimation. Therefore, the RF classifier might be another good alternative for saltmarsh wetland classification and species based biomass modelling.

***Keywords**: Saltmarsh, classification, biomass, remote-sensing, variable selection, Support Vector Machine (SVM) and Random Forest (RF).*

4.1 Background

Satellite data based image classification is one of the main focuses of the remote-sensing community, as classification results are the basis for many research questions related to environmental and ecological variables. Thematic mapping of land cover from remotely sensed data is commonly based on image classification. Lu and Weng (2007) provide a review of commonly used classification methods applied to remotely sensed data and these methods can be divided into common and advanced image classifications (Tso and Mather 1999). For example, the Maximum Likelihood Classifier (MLC) is a commonly used supervised classifier whereas ISODATA and K-Means are unsupervised methods. Advanced classification algorithms include support vector machines (SVM), random forest (RF), artificial neural networks (AN,N), and Decision tree classifier (DTC) (Adam et al. 2014; Adam et al. 2010).

The performance of these classifiers is varied with the quality of the remote sensing data either collected by the passive sensor (multispectral or hyperspectral data) or active sensor (LiDAR) data. The new multispectral sensors include, among others, RapidEye, and WorldView-2 are seen as a trade-off between benefits offered by multispectral and hyperspectral imagery (Mutanga et al. 2012). Due to the availability of fine spatial resolution and a reasonable number of spectral bands, these new sensors provide an opportunity for more detailed species classification in a complex environment. Consequently, there is a constant need for improvement of classification algorithms to deal with these new sensors and their potential applications.

One of the major problems relating to the supervised classification lies in the definition of a proper training set size for an accurate training of classifiers(Chi et al. 2008), because the collection of ground-reference data is an expensive, time-consuming and complex task. Therefore in many cases, the number of training samples is insufficient for a proper learning of classification systems. Within the supervised parametric classifiers, Maximum Likelihood Classification (MLC) is the most commonly used classifier and deliver excellent results when dealing with unimodal data. Because this classifier assumes a normal data distribution and has limitations when dealing with multi-modal input datasets.

However, in most of the cases, the normal distribution is the main violation of Remote Sensing (RS) data because the nature and causes of spatial variation in images are not always easily understandable. Hence the analysis has been limited to the empirical association between surface phenomenon and patterns in images with the implicit assumption that reality has a consistent spectral response in imagery. However, as a

consequence of the complex interplay between factors like scene complexity, scale and aggregation this assumption is often violated. Therefore, simple classifiers like MLC may reach their limits in many applications.

Non-parametric supervised classifier, such as Support Vector Machine (SVM) (Mountrakis et al. 2011), Random Forest (RF), Artificial Neural Network (ANN) (Mas and Flores, 2007) and the Classification and Regression Tree (CART), are becoming increasingly popular classifiers for remote sensing data as they do not make any assumptions regarding frequency distribution. They construct a set of classifiers and then classify new data points by taking a vote for their prediction, and, are known as ensemble classifiers. In particular, ensemble classifier, RF and SVM have received considerable attention due to several superior image-handling abilities. For example, they have the ability to handle unbalanced dataset and synthesize regression or classify functions based on discrete or continuous data sets (Breiman 2001). Moreover, they are insensitive to noise or overfitting of the model. Therefore, in recent years, remote sensing attention has moved into an ensemble classifier (Gislason et al. 2006). Ensemble classifier can be based on an individually supervised classifier or on a number of different supervised classifiers that are trained using bagging (Breiman 2001) or boosting approaches (Schapire 2003), or variations of these approaches. Machine learning takes a large collection of individually imperfect models, and, assume one-off mistakes are probably not going to be made by all models. That's how ensemble models work, they build a lot of different models, and let their outcomes be averaged or voted across the group.

Previous studies have reviewed image classification methods (Tso and Mather 1999). However, a comprehensive up-to-date review of classification approaches and techniques specifically for saltmarshes or wetland ecosystem is not yet available. Although SVM related remote sensing works have been reviewed by Mountrakis et al. (2011). New articles have been published since then, and new methods have been used for SVM. Very recently, Belgiu and Drăguţ (2016) reviewed Random Forest and its application in remote sensing. But the potentiality of ensemble classifiers for a specific ecosystem like saltmarsh and wetland ecosystem is lacking. That is why, emergence and modification of machine learning algorithms and techniques in recent years necessitate such a review, which will be highly valuable for guiding or selecting a suitable classification procedure for a specific ecosystem.

Very few studies have systematically investigated the utilization of the RF and SVM classifier for wetlands species mapping (Gislason et al. 2006; Tian et al. 2016). There has, however, been no publication to date dedicated to summarizing the use of these two machine learning classifiers in wetland species mapping and biomass model. This review will focus on this issue. Moreover, it will give a future indication for endangered saltmarsh species mapping and biomass estimation from these two algorithms.

4.2 Materials and structure of this review article

This review focuses on research articles (available in June 2017) published in nine major journals (table 4-1) of remote sensing. Google Scholar search engine (https://scholar.google.com.au/) was the base search engine where listed keywords (shown after the abstract) and synonyms of the keywords (e.g. bog /swamp for saltmarshes and estuary for wetland etc.) were used to find the best articles. A limited number of research papers relevant to the review objectives were selected from the high-quality journals (Table 4-1). Based on the selected articles, we explored only the classification of land cover, species mapping and then evaluated the potentiality for saltmarsh ecosystem.

Table 4-1: Number of articles used for this review paper

Serial No	Name of Journals	No of articles
1	Remote Sensing of Environment	10
2	ISPRS Journal of Photogrammetry and Remote Sensing	8
3	International Journal of Applied Earth Observation and Geoinformation	9
4	Remote sensing (MDPI)	9
5	Remote sensing letters (MDPI)	15
6	International Journal of Remote Sensing.	18
7	International Journal of Digital Earth.	10
8	IEEE Geoscience and Remote sensing letters	6
9	IEEE Transactions on Geoscience and Remote Sensing	11

4.2.1. Support Vector Machine

SVM is a machine learning distribution free classifier and does not encounter any over-fitting problem (Burges 1998; Cortes and Vapnik 1995). This supervised method is trained to find an optimal classification hyperplane by minimizing the upper bound of the classification error. There are two supporting hyperplanes (figure 4-1) on the boundaries of the data distribution and the data points on the edge of these hyperplanes are the support vectors of the algorithm. But the problem is all classes of an image are not linearly separable, hence, SVM is optimized to search for a non-linear hyperplane in a multidimensional feature. This transformation is implicitly performed by applying kernel functions to the original data (Keramitsoglou et al. 2006). There are two commonly used functions on remotely sensed data, non-linear polynomial and radial basis function (RBF) kernels (Huang et al. 2002; Oommen et al. 2008).

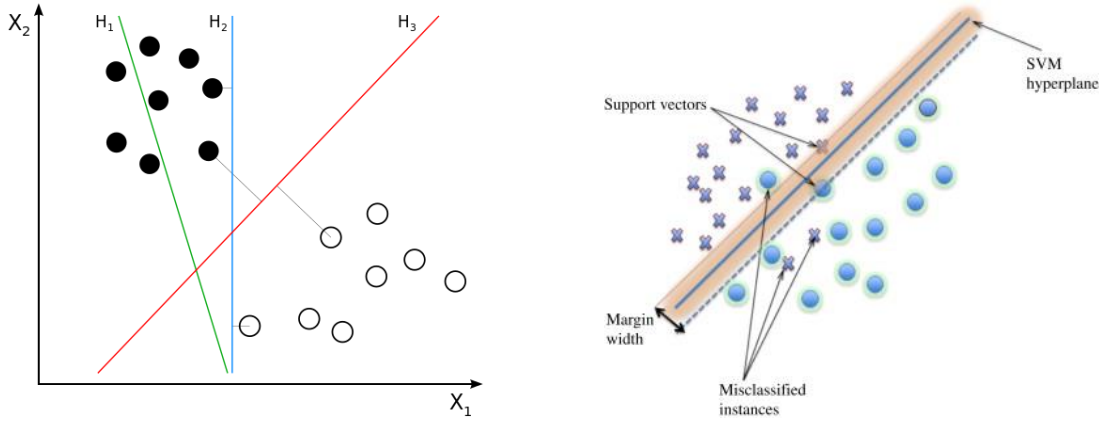

Figure 4-1: (Left) H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximum margin. (right) Linear support vector machine example. Source: (Burges, 1998; Mountrakis et al., 2011).

To solve the multiclass problem, all possible classifiers are performed using either a one-against-one (binary) or one-against-all procedure and then a voting mechanism is followed to assign the correct class (Keramitsoglou et al. 2006; Mazzoni et al. 2005). When SVM is used to solve the regression problem, it is termed as SVR. Standard SVR is the ε-SVR, where the sample points that support the "decision surface" or "hyperplane" are known as support vectors. These vectors fit the data according to the criteria of Epsilon (ε), gamma (γ) and cost (C) parameters. Here ε controlled the width of the epsilon-insensitive zone, used to fit the training data, and its value can affect the number of support vectors used to construct the regression function (Cherkassky et al. 1999). The value of ε determines the level of accuracy of the approximated function. It relies entirely on the target values in the training set. If epsilon is larger than the range of the target values, a good result cannot be expected. For example, if ε is 0, it will cause an overfitting problem. By contrast, the bigger the epsilon, the fewer support vectors are selected although bigger ε values result in more 'flat' estimates of the model (Durbha et al. 2007). The parameter cost (C) determines the balance between the model complexity (flatness) and the degree to which deviations larger than epsilon (ε) are tolerated in the optimization formulation (Durbha et al. 2007). For example, if C is too large, then the objective is to minimize the empirical risk disregard to the model complexity in the optimization. Schölkopf and Smola (2002) proposed a modification to the ε-SVR algorithm based on the difficulty in finding suitable values for the tube width ε. This modified method, called v-Support Vector Regression (v-SVR), automatically minimizes ε depending on the properties of the data (Axelsson et al. 2013). Here a new parameter (v) was introduced, in effect determining a fraction of the data points to be used as support vectors. The parameter C in the ordinary ε-SVR formulation is replaced by a parameter v which is bounded by 0 and 1. Earlier the parameter C could have taken any positive value, thus this additional bound is beneficial in implementation. The parameter v represents the lower and upper boundaries on the

number of examples that are support vectors and that lie on the wrong side of the hyperplane, respectively. However, a more detailed description on SVM theory and mathematical formulation can be found in (Cortes and Vapnik 1995; Mathur and Foody 2008a).

4.2.2. Random forest

RF is an ensemble learning technique that uses a set of Classification and Regression Trees (CARTs) to make a prediction (Breiman 2001). This algorithm creates a subset (figure 4-2) of training samples through replacement (a bagging approach). It means that the same sample can be selected several times, while other samples may not be selected at all. Usually, two-thirds of the samples (referred to as in-bag samples) are used to train the trees (*ntrees*). The remaining one third (referred to as the out-of-the-bag sample) is used in an internal cross-validation technique for estimating how well the resulting RF model performs (Breiman 2001). It means the Out-Of-Bag (OOB) sample (one-third of the total sample) is used to estimate the misclassification error (OOB error) and variable importance. At each node, a given number of input variables (*mtry*) are randomly chosen from a random subset of the features and the best split is calculated by utilizing only this subset of features (Figure 4.2). RF has recently been used as an algorithm for remote sensing image classification (Adam et al. 2014; Ozesmi and Bauer 2002). Especially, its ability to handle high dimensional and non-normally distributed data has made it an attractive and powerful option for integrating different imagery sources and ancillary data sources into image classification workflows (Kloiber et al. 2015). For a more detailed description on RF theory and its parameter optimization, the reader is directed to the study done by Tian et al. (2016), (Breiman 2001) and (Mutanga et al. 2012).

The application of the RF classifier has received increasing attention over the last decades due to the speed of processing and the excellent classification results derived from this classifier (Du et al. 2015; Rodriguez-Galiano et al. 2012). Another distinct advantage is that this classifier can be successfully used to select and rank those variables with the greatest ability to discriminate between the target classes, i.e. feature selection. Based on the application and methods used in RF, Belgiu and Drăguţ (2016) synthesized last decade work and made some future recommendation.

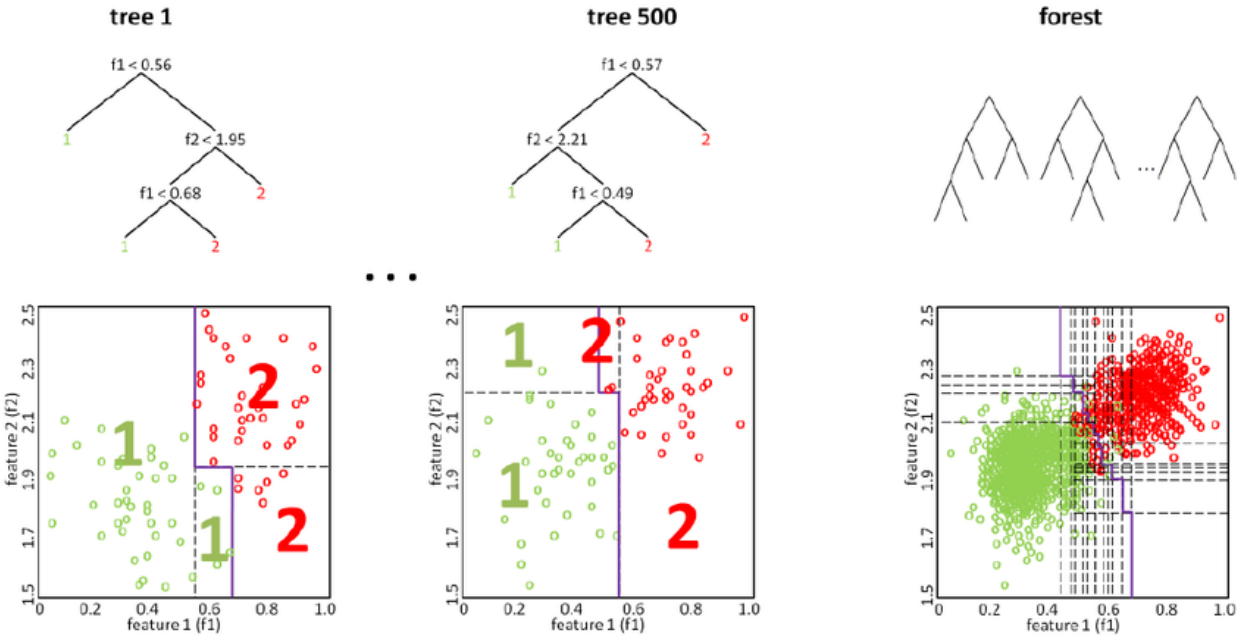

Figure 4-2 On the left and in the center, two trees of the forest are shown in detail: At each node, the feature which allows for the best class separation is chosen (with respect to the subset of features selected for that node). The corresponding partitioning of the feature space is shown below with the decision boundary plotted in purple. On the right, the decision boundary of the Random Forest is displayed. It is based on the majority votes of the individual trees (Adapted from (Hanselmann et al. 2009)).

However, based on the review (Belgiu and Drăguţ 2016) of RF, it can be concluded that there is still a research paucity on sample size allocation and mislabelling of data. Because Mellor et al. (2015) found that RF classification was relatively insensitive to mislabelled training data. Therefore that imbalanced training data can be introduced to minimize the errors in those classes that pose the greatest challenges to classifications. On the other hand, Millard and Richardson (2015) and Dalponte et al. (2013) also revealed that the RF classifier fails to cope with imbalanced training data and tends to favour the most representative classes. In another study, Jin et al. (2014) concluded that the proportionally allocated training sample design reduces the commission error of the under-represented classes and that the equally allocated training sample schema reduces the omission error of the under-represented classes. So, the impact of sampling design on RF classification results seem to be contradictory. Hence subsequent studies are required to analyse the sensitivity of RF classifier to training samples when using this classifier for remote sensing data classification.

4.3 Overview of Support Vector Machine application for vegetation mapping and land cover classification

It is already mentioned that most of the works of SVM have been summarized by Mountrakis et al. (2011) in their review work. In this article, we are going to emphasise the works based on a land cover mapping that has been done since 2002 and more focus on the work that has been done after 2011. Figure 4-3 is showing the development of SVM since 2002. It is clear that most of the works on SVM application has been done between 2012 and 2013. However, to review the overall developments of key findings (table 4-2) and methodological development (table 4-3) of SVM, we reviewed all of the works where SVM has been used for land cover mapping and wetland monitoring.

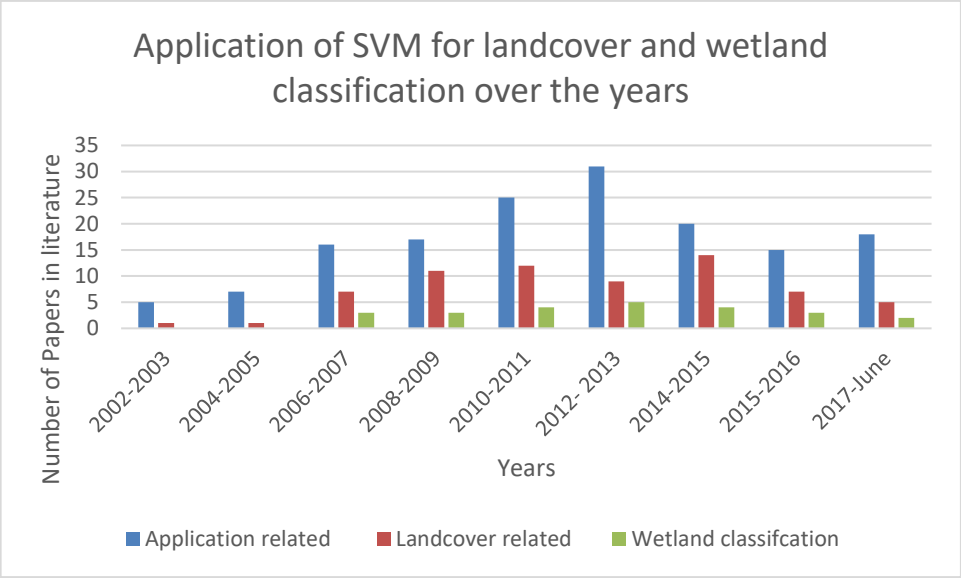

Figure 4-3: Number of papers on SVM and its application in land cover classification published since 2002.

Table 4-2: Application of SVM in land cover classification and the key findings.

Authors, data and area of interest	Key findings
Gualtieri and Cromp (1998) used Hyperspectral AVIRIS imagery for Vegetation classification.	SVM was superior over prior classifiers developed on the same dataset
Camps-Valls et al. (2004) used Hymap hyperspectral data for Crop classification.	SVMs were not sensitive to training sample size
Keramitsoglou et al. (2006) used IKONOS imagery for vegetation mapping.	SVM training time was considerably lower suggesting improved applicability for vegetation mapping.
Knorn et al. (2009) used Landsat data for binary forest classification	Chain classification accuracy, which proved accurate even for lengthy sequences (e.g., six images).
Lardeux et al. (2009) used SAR data for tropical vegetation classification.	SVMs resulted in about 20% higher classification accuracy than the Wishart classification approach.
Dalponte et al. (2008) used hyperspectral data for forest species classification.	SVMs outperformed Gaussian maximum likelihood classification and k-NN technique.
Dalponte et al. (2013) used boreal forest with HySpex hyperspectral data.	There was no significant difference between the performance of SVM and RF classifier.
Su and Huang (2009) evaluated different linkage techniques on semi-arid vegetation mapping	A reduced dataset of approximately 20% of the original size could provide comparable classification accuracy.
Shao and Lunetta (2012) used MODIS data for land cover classification in Albemarle-Pamlico Estuarine System (APES) in North Carolina and Virginia, USA	Sample size, sample variability and landscape homogeneity analysis based on three classifiers Neural Network, SVM and CART. SVM outperformed (77-80%) the CART (62-73%) and NN (67-76%).
Zhang et al. (2015) compared traditional single optimized parameters (SOP) with multi-parameter SVM.	Single optimized parameter (SOP) and a novel multi-parameter (MP) both are sensitive to landscape to be classified.
Yang et al. (2016) Wetland mapping	SVM provided higher accuracy
Fang et al. (2016) evaluated dynamic change analysis of wetland with time series data.	Decision tree classifier (DTC) works better compare to the MLC and SVM

4.3.1 Evaluation SVM classifier and its sensitivity

In early works, the main emphasis was to compare the accuracy with another parametric classifier. For example, Keramitsoglou et al. (2006), Su and Huang (2009), Dalponte et al. (2008), Lardeux et al. (2009) and (Knorn et al. 2009) they all focused on the improvement of overall classification accuracy. However, the main appealing feature of SVM is 'small training sample' or reduced training data.

Sample size

Mathur and Foody (2008b) showed that SVM may be used to derive very accurate classifications (90.66%) using a smaller, intelligently defined training set. This result was statistically insignificant (Z =1.50) when it was compared with the traditional training sampling approach (92.00%). This 1.34% decrease in accuracy was achieved with a decrease in training set size from 450 to 130 pixels that were a huge savings of time and cost. Another experiment that revealed the ability to improve classification accuracy from small training sample is the study of Ghoggali et al. (2009). Later, Su and Huang (2009) investigated training data reduction using a hierarchical clustering analysis and Multiangle Imaging Spectro-radiometer satellite data on a vegetation classification problem. They proved that a two-thirds reduction of the dataset size was possible without significant accuracy degradation in SVM method.

However, Dalponte et al. (2013) argued that SVM can be considered a complex classifier, both theoretically and from a usability viewpoint. Due to its nonlinear form with RBF kernel function, it requires the optimization of two parameters. This optimization process increases the computational time with respect to the other two classifiers (Random forest and Gaussian Maximum Likelihood). But the main advantage of SVM is that it provides good classification accuracies, even for minority classes (Dalponte et al. 2013). As the recent trend of SVM turned to focus on methodological approach, we reviewed (table 4-3) different modification, kernel selection, parameter optimization and customization that have been done on SVM based on the classification aim and study site complexity of a landscape.

Table 4-3: Application of different methods in SVM platform to solve classification problem

Method	Authors	Application and Description
A recursive procedure	Mantero et al. (2005)	They used Bayesian minimum error decision rule to generate prior probability estimates for known and unknown classes.
Transductive inference theory	Bruzzone et al. (2006)	Define the separating hyperplane according to the process that integrates the unlabelled samples together with the training samples.
Mixed pixel training samples	Foody and Mathur (2006)	Focused on mixed pixels training samples over more tedious, conventional pure pixels acquisition. Although there was no significant difference between mixed pixels spectral responses and the conventional approach, SVM based mixed pixels approach was easier and cheaper.
Data reduction	Foody et al. (2006)	They evaluated four dataset reduction methods for a one-class problem (cotton vs others) using SVMs. They concluded that significant data reduction (approximately 90%) was possible with minimum information loss.
	Sahoo et al. (2007)	They used SVM that showed additional robustness to small data samples in a geological classification.
	(Su 2009)	Two-thirds reduction of the dataset size was possible without significant data.
Primal SVM	Chi et al. (2008)	The primal SVM formulation makes it possible to optimize directly on the primal representation and therefore limits the number of samples. Primal SVM yielded competitive accuracy values as the state-of-art alternative algorithms trained on larger datasets.
Bootstrapped SVM	Castillo et al. (2008)	The training strategy adapted in the bootstrapped SVM is that an incorrectly classified training sample in a given learning step is removed. Later it was re-assigned with a correct label and re-introduced into the training set in the subsequent training cycles. Only 0.05% of the total number of training pixels were needed to achieve about the same accuracy level as the standard SVM.
Ensemble methods	Pal (2008)	Two popular integration techniques such as boosting (alternating observation weight) and bagging (alternating observations) were tested using Landsat ETM+ data for an agricultural classification.
	Tan et al. (2007)	Combining entropy decomposition and SVM for classification.
	Tuia et al. (2009)	Combined morphological filters and SVMs to conduct land use classification using high spatial resolution Quickbird panchromatic images.
Scene segmentation with SVM	Li et al. (2010)	A scene segmentation algorithm was integrated with the SVM object classifier that provided a better performance with Quickbird high-resolution imagery.

Parameter settings

One of the major challenges regarding the application of SVMs is the choice of kernels. Although there are many options for the kernel, some of the kernel functions may not provide optimal SVM configuration for remote sensing applications. Linear, polynomial and a Gaussian radial basis function (RBF) are generally used as the kernel function. However, empirical evidence indicates that different kernels applied to the SVM-based classification of satellite image data produce different results (Zhu and Blumberg 2002). The input data are first transformed into a high-dimensional feature space using a kernel function. By transforming the data using a non-linear kernel function, it is possible to map non-linear relations using a linear model in the new feature space. The algorithm takes both calibration error and model complexity into account when constructing the SVM model, and their relationship is defined by setting regression parameters cost (C) and gamma.

Feature selection on SVM

Feature selection addresses the problem of finding the most compact and informative set of features, to improve the efficiency or data storage and processing. The feature is synonymous with input variable or attribute used in machine learning model (i.e. a number of bands of satellite data). There are three variable selection methods- filter, embedded and wrapper method (Mehmood et al. 2012; Saeys et al. 2007).

Detailed information of these three methods is available in the studies of Guyon and Elisseeff (2003) and Saeys et al. (2007). SVM is an effective classification method, but it does not directly obtain the feature importance. There is a limited number of works found on variable selection in SVM classifiers. Fisher's criterion, output Pearson correlation coefficients, and mutual information are common filter method that used in variable selection for SVM (Guyon and Elisseeff 2003). Support vector machine recursive feature elimination (SVM-RFE) is an example of an embedded feature selection method (Guyon et al. 2002) that has both linear and nonlinear version according to the type of kernel function. SVM-RFE uses criteria derived from the coefficients in the SVM model to assess features, and recursively removes features that have small criteria (Tang et al. 2016). SVM-RFE is able to make full use of the training data since the data don't divide into the test set and training set (Tang et al. 2007). However, applying different feature selection methods does not necessarily result in significant differences in SVM classification performance (Pal and Foody 2010). Fassnacht et al.,(2012) reported the feature selection method that is independent of the SVM and can even result in better accuracy, e.g. SVM-RFE (Bazi and Melgani 2006; Pal and Foody 2010). Besides SVM-RFE method, direct objective optimization with k1-norm regularization (Be et al., 2003), and a correlation bias reduction technique (Tang et al. 2016; Yan and Zhang 2015) is proposed to further improve the performance of feature selection and classification results.

Pal (2006) used the SVM method for feature selection. In his study, SVM-based methods combined with GA were compared with the random forest feature selection method in land cover classification problems with hyperspectral data and small benefits were identified. In another study, Archibald and Fann (2007) introduced an embedded-feature-selection (EFS) algorithm within the SVM classification approach. They compared EFS with RFE and showed that major advantage of EFS is the significant reduction in computational time and the data-driven knowledge discovery of important bands in classification. Zhang and Ma (2009) used SVM approaches for feature selection but they implemented a modified recursive SVM approach to classify hyperspectral AVIRIS data and found slightly better results. However, their method has higher computational demands compared with others. Further, the cost and gamma parameters selection in conjunction with feature selection when using SVM-RFE is still an issue (Guyon et al. 2002).

4.3.2 Biomass modelling using Support Vector Machine Regression (SVR)

There are multiple methods (linear regression, stepwise multiple linear regression, linear mixed ordinary least square regression, partial least squares regression) used to estimate biomass, which is varied in their assumptions and complexity to develop biomass model. The advantage of linear mixed effects regression and geographically weighted regression is that these sophisticated regression techniques take into account bias and the correlation of predictor variables rather than the somewhat rigid assumptions of ordinary least squares regression (Powell et al. 2010; Salas et al. 2010). Zhao et al. (2009) also noted that regression models are built to output biomass at a specific plot size and changing this plot size may affect the accuracy of results. Because population assumptions of regression models do not represent the heterogeneity of forest stands (Robinson et al. 2013). To reduce the effects of regression assumptions on plot scale biomass estimation, machine learning techniques such as Support Vector Machine (SVM) and Random Forest (RF) may be used (Breidenbach et al., 2010; Vauhkonen et al., 2010). Because SVM and RF are non-parametric in the sense that no parametric model or distribution is assumed (Neumann et al. 2012; Robinson et al. 2013).

Machine learning classification technique, SVM is also adapted to regression problem. However, very few studies have been found where this method has been used to estimate biomass. The AGB estimation performance based on SVM method are summarized (Table 4-4) below:

Table 4-4: Summary of SVM application for biomass estimation

LC	Authors	Major contribution
Wheat (Field condition)	Montes et al. (2011)	Support Vector Machine Regression (SVMR) with A Gaussian Radial Basis Function (RBF) provided the lowest error of prediction (0.11) and highest R^2 (=0.97) in compare with Partial Least Square Regression (PLSR) for Maize biomass under field conditions.
Conifer and deciduous forest	Chen and Hay (2011)	SVR models achieved better performance for estimating canopy height (R 2= 0.81, RMSE = 4.0) and above ground biomass (R^2 = 0.76; RMSE = 63.1 Mg/ha) than multiple regression.
Boreal forest	Neumann et al. (2012)	SVM and RF did not succeed in improving the cross-validated results because these methods are partly overfitted the data due to the noisy nature of the radar observables and an insufficient number of training samples.
Heiberg Memorial Forest, Tully, NY, USA	Gleason and Im (2012)	Among the four different models (Linear Mixed-effects, Random Forest, SVM and Cubist) ,SVR produced the most accurate biomass model at the plot level, although all models provided similar results at individual tree level.
Peat Swamp forest, Indonesia	Englhart et al. (2012)	Multivariate Linear Regression (MLR), Artificial Neural network (ANN) and Support Vector Machine (SVM) were examined for their performance to retrieve AGB from multi-frequency SAR data. The SVR modelled AGB was more accurate than ANN modelled AGB in terms of independent validation.
Grasslands	Marabel and Alvarez-Taboada (2013)	Spectroscopic determination of aboveground biomass was done using spectral transformation , SVM and Partial Least Square Regression (PLSR). It was proved that SVM outperformed PLSR data when no transformation was applied to the reflectance data.

4.3.3 Future directions: Applications of SVM for saltmarsh vegetation mapping and biomass estimation

Selection of SVM key parameters i.e. the kernel functions is one of the important limitations to SVM methodologies application for remote sensing data. Because, a small value for the kernel width parameter may lead to overfitting, while large kernel width values may lead to over-smoothing. However, this problem is a general drawback of kernel-based approaches (e.g., radial basis function neural networks). Settings of the parameter value cost (denoted by C), which controls the trade-off between maximizing

the margin and minimizing the training error, is another important consideration in SVM application. Another limitation is SVM approaches map input data to higher dimensional spaces. As dimensionality increases in addition to potential separability of patterns SVMs exhibit outlier behaviour and increased computational demands. This is a critical drawback for hyperspectral analysis where the dimensionality of the original data is high and kernel mapping is more vulnerable to dimensionality problems. SVM follows 'one-against the rest' strategy for multi-class classification problem and can be problematic as it may result in unclassified instances of data. Therefore lower accuracy (Pal and Mather 2005) may be an issue. Sometimes due to the limited precision of image acquisition instruments, and atmospheric and topographic distortions cause measurement errors in remote sensing data. SVMs are not optimized to deal with this inherent problem of noisy data (outlier effects). In addition, the performance of an SVM can decrease with a relatively small number of mislabelled examples.

Although there are some pitfalls, however, there is significant scope for extension of SVMs to address these pitfalls. As an example, Foody (2008) evaluated a relevance vector machine approach (RVMs) as a way to address the need to define the cost parameter C. RVMs have several advantages including automatic estimation of parameters and the arbitrary kernel functions that are the main challenges of SVM. In addition, RVMs allow for sub-pixels classification (fuzzy) of data making it possible to have a probabilistic output. More investigations into the potential of some of the relatively untapped lower level noise reduction techniques such as morphological image processing might be a remedy to the problem of noising. Regarding the novelty of SVM for wetland classification, Zhang et al. (2015) proposed a multi-parameter (MP-SVM) algorithm that divides the training set into several subsets which are subsequently combined. Based on these combinations, sub-classifiers are constructed using their own optimum parameters, providing votes for each pixel with which to construct the final output. Based on the complexity of landscape they found a different result for MP-SVM and SOP-SVM (single optimized parameter). For the high and moderate complex landscape, MP-SVM provided better performance (OA= 82.19 %, Kappa= 0.80) than SOP, however, for low complex landscape there was no significant difference between them. So this is a future indication to consider landscape complexity before making any customization on SVM. On the other hand, Fang et al. (2016) showed that Decision tree classifier (DTC) performs well in compared to SVM and MLC when they mapped a wetland with time series data. It gives an indication that still there is an opportunity to work with SVM for saltmarsh ecosystem with another non-parametric classifier.

4.4 Overview of Random Forest application based on vegetation mapping and land cover classification

It is already mentioned that RF is fairly new in the remote sensing image analysis technique. However, this ensemble machine learning algorithms already got a special

attention from the researcher due to its unique features. Various ensemble classification methods have been used in recent years including boosting and bagging. Boosting is based on sample re-weighting and bagging is bootstrapping of the sample. RF uses bagging, or bootstrap aggregating to form an ensemble of classification and regression tree (CART)-like classifier (Gislason et al. 2006). In addition, this method searches only a random subset of the variables for a split at each CART node, in order to minimize the correlation between the classifiers in the ensemble (Breiman 2001).That is why RF is not sensitive to noise or overfitting. Moreover, the RF does not require any assumptions about the relationships between explanatory and response variables and are well suited for analysing complex non-linear and possibly hierarchical interactions in large data sets (Olden et al. 2008) that are often a requirement to solve a complex classification problem. The Random Forest classifier has successfully been used to map land use and land cover classes (Colditz 2015; Haas and Ban 2014; Stefanski et al. 2013), tree canopy cover and biomass mapping (Karlson et al. 2015) and tree health (Wang et al. 2015). Räsänen et al. (2013) used Worldview-2 imagery to map boreal forest, Waske and Braun (2009) used Synthetic Aperture Radar (SAR) and Uhlmann and Kiranyaz (2014) used PolSAR data for land cover classification. Figure 4-4 highlighted the chronological development of RF over the years and Table 4-5 summarizes the application of the RF tool for different vegetation species classification. It is already mentioned that the application and methods of RF have been summarized by Belgiu and Drăguţ (2016). In our study, we are going to emphasize the overall development of RF in respect of land cover classification only. Then we provided a special attention to the wetland ecosystem and the application of RF for wetland classification and monitoring.

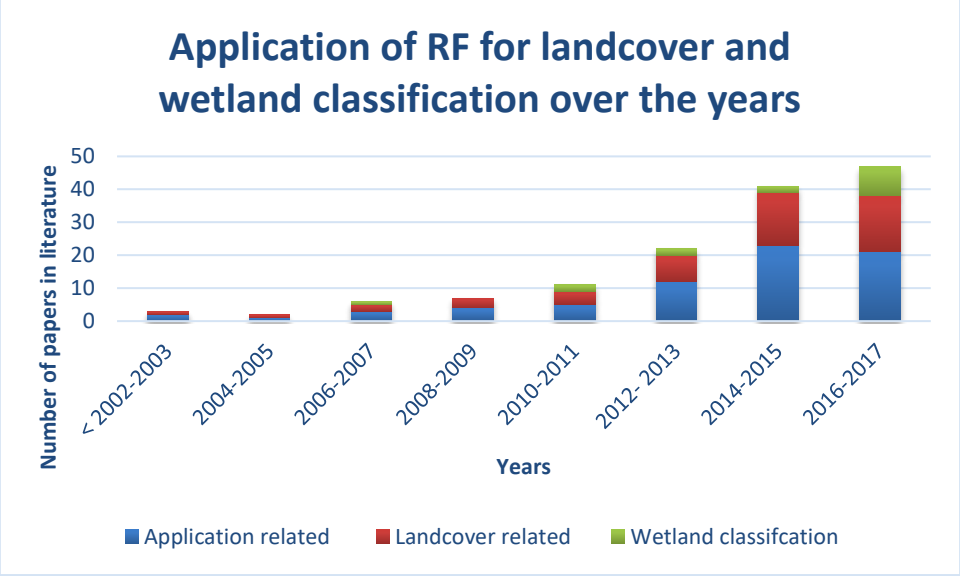

Figure 4-4: Growth of RF and its application in land cover classification over the years.

Table 4-5: Summary of RF application in different vegetation species a classification and mapping.

Authors	Field of study	Description
Ham et al. (2005)	Forest Classification with Hyperspectral data	Two modified approaches are proposed based on the concept of RF and both of them provided a superior result in compare to the original frame.
Lawrence et al. (2006)	Invasive species classification with hyperspectral data	Beiman Cutlar Classification (BCC) from RF can achieve substantial improvements in accuracy over single classification trees.
Cutler et al. (2007)	Ecology: Species classification	Compared RF accuracy with other techniques to classify various species. RF proved to be superior to others
Chen and Ho (2008)	Eco tops	Comparison of RF with Adaboost and Neural network. RF outperformed neural network.
Mansour et al. (2012)	Rangeland degradation	Application in Hyperspectral data
Rodriguez-Galiano et al. (2012)	Land cover classification	Compare RF with a single decision tree
Du et al. (2015)	Synthetic Apperture Radar (SAR) image classifcation	RandomForest and Rotation forest comparison
(Karlson et al. 2015)	Landsat 8 for tree canopy cover	Tree Canopy Cover was identified from RF with the coefficient of determination 0.77 (RMSE = 8.9%).
Shaohong et al., (2016)	Wetland mapping	Comparison of SVM, RF and NN. RF outperformed SVM and NN.
Fu et al. (2017)	Wetland mapping	Pixel and object-based classification using RF through wavelet principal component analysis (PCA).
Mahdianpari et al., (2017)	Wetland classification	A novel hierarchical object-based RF classification is proposed.
Franklin and Ahmed (2017)	Wetland classification	RE method has been used for object and pixel-based classification comparison and integration.
Amani et al. (2017)	Wetland classification	Integration of SAR and Optical through OBIA.

In respect of vegetation mapping, Ham et al. (2005) investigated two approaches based on the concept of random forests classifiers when the quantity of training data is limited for hyperspectral data. Proposed two approaches were superior to those from the original basis, binary hierarchical classifier (BHC) algorithm and a random subspace extension of the BHC. Later, Cutler et al. (2007) highlighted the major advantages of RF in compare to the other four classifiers. According to Gislason et al. (2006) and Cutler et al. (2007), RF has (1) very high classification accuracy; (2) a novel method of determining variable importance; (3) ability to model complex interactions among predictor variables; (4) flexibility to perform several types of statistical data analysis, including regression, classification, survival analysis, and unsupervised learning; and (5) an algorithm for imputing missing values (6) computationally much lighter. They observed high classification accuracy in all applications of RF when comparing RF to other common classification methods for invasive plant species. Similarly, Chen and Ho (2008) compared RF with Adaboost (REF) and showed that Adaboost and Random Forest attain almost the same overall accuracy (close to 70%) with less than 1% difference, and both outperform a neural network classifier (63.7%). Random Forest, however, is faster in training and more stable. In a very recent study, Kalson et al (2015) achieved R2= 0.77 (rmse = 8.9 %) from a freely available Landsat 8 for tree canopy classification.

4.4.1 Evaluation RF classifier and its sensitivity

Hughes (1968) conducted a statistical analysis and proved that the accuracy of a supervised classifier depends on the number of training samples or the data dimension. Later, this curse of dimensionality is also known as the Hughes effect or the Hughes phenomenon Hughes (1968). Lavergne and Patilea (2008) proposed a general nonparametric method trying to avoid or reduce the Hughes effect. Non-parametric classifiers work to accommodate the increasing number of data dimensions to mitigate the Hughes phenomenon. Therefore, RF is also a supervised classifier and, therefore, it is important to evaluate the sensitivity of RF classification, how it responds to the size of the training samples, data dimension and sampling design.

Sampling design

Colditz (2015) explored the sensitivity of the RF classifier to different sampling designs such as random, area proportional sample number for each class and an equal number of samples to each class. This land cover classification study revealed that area-proportional allocation of training samples per class provided the best classification accuracy. Because the classes that occupy large areas require more samples than those that occupy small

areas Colditz (2015) for random forest classification. On the other hand, Mellor et al. (2015) found that RF classification was relatively insensitive to mislabelled (Forest as non-forest or vice versa) training data and that imbalanced training data can be introduced to reduce the errors in those classes that pose the greatest challenges to the forest classifications from Landsat data. Other land cover classification studies (Rodriguez-Galiano et al. 2012; Rogan et al. 2008) also demonstrate that RF is resistant to the mislabelling of the training data.

Size of the training samples

The sizes of the training samples sets have been found to influence the performance of the RF classifier for Landcover classification (Colditz 2015). Although Rodriguez-Galiano et al. (2012) showed that RF has low sensitivity to the training set size reduction, investigations to date have reported that the sizes of the training samples have high sensitivity to the performance of RF classification (Colditz 2015). Ma et al. (2017) found that RF is less sensitive to the effect of data dimensionality compared to the SVM, even though a small training set size is used. Li et al., (2016) proved that either RF or SVM could be used with limited training samples sizes.

Spatial autocorrelation

Ham et al. (2005) investigate two approaches based on the concept of random forests of classifiers implemented within a binary hierarchical multi-classifier system. They found that RF methods yielded superior results for both test and spatially disjoint test data two study sites, thereby indicating improved generalization to extended areas. Similarly, Rogan et al. (2008);Rodriguez-Galiano et al. (2012) and Colditz (2015) also found that RF is insensitive to spatial autocorrelation. On the other hand, Millard and Richardson (2015) found that RF classifier is sensitive to the spatial autocorrelation of the training classes and to the proportions of the different classes within the training samples.

Feature space optimization

Random forest classifier can handle high dimensional data based on feature selection methods. This classifier has been tested as a feature space optimization technique with satisfactory results in numerous applications such as the mapping of land cover classes classification (Chan and Paelinckx 2008; Demarchi et al. 2014), tree species classification(Cavallaro et al. 2016), invasive plants (Lawrence et al. 2006; Peerbhay et al. 2016), and grass species (Mansour et al. 2012). A number of studies have investigated various alternative methods for feature space optimization (Karlson et al. 2015; Millard and Richardson 2015; Stumpf and Kerle 2011).

There are a number of feature selection methods used in combination with RF classification, such as filter methods, embedded methods, and the wrapper methods(Saeys et al. 2007; Weinmann et al. 2015). Table 4-6 summarizes the feature selection methods that are used in combination with RF classification.

Table 4-6: Feature selection methods in RF platform for classification problem.

Filter method	
Model name	Application and description
Relief and (MNA)	Yu et al. (2016) used two filter methods, Relief and max-min-associated (MNA) to explore the potential of feature selection in global land cover mapping.
Wrapper method	
Backward feature	Díaz-Uriarte and Alvarez de Andrés (2006) proposed an iterative backward feature elimination procedure to reduce the number of less relevant variables.
Forward feature	Mansour et al. (2012) used this method to identify grass species form hyperspectral data; Dalponte et al. (2013) used Sequential Forward floating (SFS) for tree species classification from Hyperspectral data
Embedded methods	
RF and Out of Bag (OOB) error	Kursa and Rudnicki (2010) used RF algorithm to reduce noisy and highly correlated variables derived from LIDAR and Hyperspectral data to model plant composition and diversity in forested areas; Adelabu et al. (2014)used Out of Bag (OOB) based feature space optimization yielded better results in comparison to ANOVA in a case of leaf defoliation analysis; Millard and Richardson (2015) applied Spearman's rank order correlation to determine pair-wise correlations between 15 important variables identified by the RF classifier (out of 28 variables).
Mean Decrease Accuracy	Franklin and Ahmed (2017) used RF variable importance that is the mean decrease in permutation accuracy with each variable considered in turn.

4.4.2 Evaluation of RF feature selection methods

Filter methods (Univariate and multivariate methods): Univariate filter methods are fast and independent and multivariate filter methods identify the most relevant features independently of a specific classifier (Saeys et al. 2007). Euclidean distance, Principal Component Analysis (PCA), Chi-square test, Independent Component Analysis (ICA), Correlation Based Feature Selection (CBFS), Minimum Noise, Fraction (MNF) analysis (Zhang, 2014), Wilks's Lambda stepwise discriminant analysis (Thenkabail et al. 2013),

linear discriminant analysis with stepwise feature selection, Markov Blanket filter (MDF) and Spearman's rank order filter are a few examples of filter methods that can be used to remove redundant, noisy and irrelevant variables. A very few studies have been found in literature where filter methods are used in combination with an RF classifier for land cover mapping. For example, (Yu et al. 2016) explored the potential of feature selection in global land cover mapping through two filter methods, Relief and max-min-associated (MNA) in RF classifier and found no significant changes in accuracy after the reduction of variables from 63 based on the mentioned methods. One possible reason of this limited use of filter method is that no significant differences are gained between the RF classification results obtained by applying these filtering methods and those obtained using the feature selection methods embedded in the RF classifier (Dalponte et al. 2008; Millard and Richardson 2013).

Embedded methods: This approach fully integrates feature selection and classifier design together. For example, decision trees, weighted naïve bayes (Duda et al. 2012) and feature selection using the weight vector of SVM (Guyon et al. 2002; Weston et al. 2003) are common embedded methods (Saeys et al. 2007) of feature selection. Löw et al. (2013) used different sizes feature subspaces in SVM classification where the dimensionality of the input data set varied by incrementally adding features in the order suggested by the random forest by means of the OBB mean decrease of accuracy (MDA). This MDA approach embedded in the RF classifier is frequently used to pre-select the most important variables (Karlson et al. 2015; Li et al. 2014). Díaz-Uriarte and Alvarez de Andrés (2006) proposed an iterative backward feature elimination procedure to reduce the number of less relevant variables. This procedure has been incorporated within the R software and known as "varSelRF" package. This package has been successfully used to classify landslide objects from high resolution optical data such as Quick Bird, IKONOS, GeoEye-1 and aerial photographs (Stumpf and Kerle, 2011) and reforested landslides from LiDAR data(Chen and Hay 2011; Li et al. 2014), and grass species from hyperspectral data(Mansour et al. 2012). Recently, Li et al. (2017) compared two feature selection methods, varSelRF and wrapper Baruta packages with three pre-filtering methods (PCA, ICA, and MNF) using RapidEye Spectral bands, several vegetation indices and SVM algorithms. According to their finding, all the five methods could improve classification accuracy, but the only varSelRF achieved significant improvement and outperformed the other methods. Zhang et al. (2016) found that embedded method allowed for assessment of the importance of variables by means of Gini index and the OOB subset. RF-based feature selection method selected 29 variables out of 114 variables (multi-temporal spectral bands, spectral indices and textural features) as important inputs and provided an overall accuracy of 89% and a Kappa statistics of 0.8522.

Wrapper methods (Deterministic and Randomized): Deterministic and randomized are two types of wrapper methods and both of them suffer from the risk of model overfitting (Sayes et al., 2007). Commonly used example of wrapper methods is a sequential forward selection (SFS) and Sequential Backward Elimination (SBE). The Boruta wrapper feature

selection method has been successfully used to reduce noisy and highly correlated variables derived from LiDAR and hyperspectral data, in order to model plant compositions and diversity in forested areas (Leutner et al. 2012). Recently evaluated feature selection methods for object-based land cover mapping using RF and SVM methods. They used five feature -importance -evaluation methods and three feature subset evaluation methods. Based on the evaluation of five feature-importance-evaluation methods, Ma et al. (2017) found that RF outperformed the SVM classifier and, the classification accuracy was relatively stable with the variation of features when small training set sizes were used. Therefore, they once again proved that the RF classifier is insensitive to the number of features, even for a small training sample size. However, when they evaluated the wrapper method with other feature selection methods, it did not retain the superiority for object-based classification although it was claimed for pixel-based classification for Hyperspectral data (Chan and Paelinckx 2008; Huang and Zhang 2013; Kohavi and John 1997).

However, an integration of embedded and wrapper feature selection methods was reported by Chan and Paelinckx (2008) and Li et al. (2017). Chan and Paelinckx (2008) examined the relevance of hyperspectral bands (HyMap data) for identifying land cover classes by applying the OOB error and the wrapper method. (Li et al. 2017) used the combination of embedded and wrapper method to evaluate the performance of land cove mapping.

4.4.3 Application of Random Forest for wetland/saltmarshes classification

van Beijma et al. (2014)used multisource data for saltmarsh vegetation mapping using multisource data and found that RF models provided higher classification accuracy from multisource data in comparison to either SAR or optical data alone. Very recently, Shaohong et al., (2016) compared three classifiers for saltmarshes mapping and RF classification stood out with the highest classification accuracy, showing its effectiveness in handling interference of soil background in the sparsely vegetated arid areas. However, in comparison with the other classes of the wetland, Shaolong et al., (2016) found relatively lower accuracy (varied from 85% to 90%) for *Phragmites australis*, *Tamarix Chinensis* and *Echinochloa crusgall*i. Based on RF platform, Fu et al. (2017) and Franklin and Ahmed (2017) found improved overall accuracy using object-based image analysis. Similarly, Amani et al. (2017)also used object-based image analysis in RF using the integration of aerial imagery, SAR and optical satellite data and achieved up to 91% overall accuracy for wetland classification. Mahdianpari et al. (2017) used a novel hierarchical object-based RF for discriminating between different classes in the wetland. Based on this hierarchical approach, they achieved an overall accuracy of up to 94% using ALOS-2 L band, RADARSAT-2 C band and TerraSAR-X imagery. So based on the application of RF on the work of saltmarshes/ wetland, it may be summarized that most of the RF application on wetland is data dependent. Different authors tested several data sources, mostly SAR or fusion data to test the feasibility of RF. Based on our knowledge, methodological

approach is very rare for RF in wetland application. Although Stumpf and Kerle (2011), Chan et al (2012) and Du et al (2015) worked with limited data for landslides, health land and urban area mapping respectively sample size behaviour with RF is rarely practised for wetland ecosystem.

4.4.4 Biomass modelling using random forest regression

Traditional regression models have limitation to adequately capture the relationship and the spatiotemporal variability of the quantity (Kaheil et al. 2008). Moreover, multicollinearity is an important issue for the multiple regression model, especially when highly related variables (i.e bands of RS data) are selected as a predictor. Multicollinearity (also collinearity) is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, and Variance Inflation Factor (VIF) is used to detect multicollinearity before the process of multiple regression. RF works on both 'bagging' and 'boosting' algorithms (i.e.bootstrap aggregating) that aim to reduce the complexity of models that overfit the training data. Bagging techniques can be used to reduce the variance in model predictions where numerous replicates of the original data set are created using random selection with replacement. Each derivative data set is then used to construct a new model and the models are gathered together into an ensemble. Clearly, the mean is more stable and there is less overfit. Therefore, RF algorithm yields an ensemble that can achieve both low bias and low variance from averaging over a large ensemble of low-bias, high-variance but low correlation trees (Breiman 2001; Díaz-Uriarte and De Andres 2006b). RF is originated from Classification and Regression tree (CART)(Breiman 2001). However, decision trees suffer when features are correlated since they choose a single feature to maximize information gain at each step instead of a combination of features that also makes them unstable to small perturbations in the dataset. This instability is exploited in a very robust way in RF by building bagged tree ensembles. The multicollinearity problem is alleviated since a random subset of features is chosen for each tree in a random forest (Díaz-Uriarte and De Andres 2006a; Liaw and Wiener 2002a). In addition, it can handle thousands of input variables without variable deletion to fit a regression model although it has fewer parameters compared with that of other machine learning algorithms (e.g SVR).

Based on the literature, only a few studies have reported the use of the RF regression in remote sensing applications. For example, the RF algorithm has been successfully employed in forest biomass (Gleason and Im., 2012; Dube and Mutanga., 2015), wetland species biomass (Mutanga et al. 2012), to map biomass using Landsat temporal data (Frazier et al., 2014), and to map tree canopy cover and biomass using uni-temporal and multi-temporal Landsat. RF has also been used in precision agriculture where crop biomass used as an indicator of vegetation development and health. For example, Wang et al., (2016) used RF for wheat biomass and compared it with the other two machine learning algorithm, SVR and ANN and found that RF performed better than SVR. The summary of RF application and its outcome for biomass estimation are presented in table 4.7:

Table 4-7: Summary of RF application for biomass estimation

Land cover	Authors and their contributions
Forest	Frazier et al. (2014) investigated the importance of types of temporal Landsat trajectory matrices in the estimation of boreal forest biomass using RF algorithm.
	Mascaro et al. (2014) evaluated the performance of Random Forest in upscaling airborne LiDAR (Light Detection and Ranging)-based carbon estimates compared to the stratification approach over a 16-million hectare focal area of the Western Amazon.
	Tanase et al. (2014) used airborne Polarimetric L-band Imaging Synthetic aperture radar (PLIS) to the impact of high revisit cycle and full polarimetric acquisitions on biomass retrieval by means of backscatter-based multitemporal methods.
	Karlson et al. (2015) evaluated the utility of Landsat 8 for mapping tree canopy cover and biomass in a woodland landscape.
	Dube and Mutanga (2015): Although the objectives of their study were not to compare algorithms, the study has shown that Stochastic Gradient Boosting (SGB) outperformed the Random Forest (RF) algorithm in all aboveground biomass estimation stages. They compared Landsat 8 OLI with Landsat 7 data.
	Gleason and Im (2012) evaluated machine learning approaches –RF, SVR and Cubist regression trees for forest biomass estimation using Airborne LiDAR data.
Agriculture	Wang et al. (2016) compared RF with other two machine learning classifiers (SVM and ANN) to estimate wheat biomass. RF outperformed ANN and SVM.
Wetland	Mutanga et al. (2012) showed that the performance of non-linear RF predictive model of the selective NDVIs was better (R^2= 0.76, rmse= 0.441) compared to stepwise multiple regression (R^2 = 0.69, rmse = 0.5465) for wetland biomass.

But for wetland biomass estimation we found only one studies that used RF algorithm (Mutanga et al, 2012). So, there is still a further huge scope to work on it. One possible reason is that most of the vegetation indices used for biomass estimation are more or less similar and correlated (References). Cutler et al., (2007) demonstrated that RF is not sensitive to collinearity. This is very valuable in modelling, especially for a complex, non-linear system because it is commonly difficult to decide which variable to remove when two (or more) variables correlate with each other.

4.4.5 Future direction: application of random forest for saltmarsh mapping and biomass modelling

To evaluate the classification accuracy of RS data, parametric classifiers need additional effort and time to collect validation samples. Lawrence et al. (2006) and Zhong et al. (2014) found a solution and reported that the OOB error in the RF method could be used as a reliable measure of classification accuracy. However, this assertion needs to be further tested using a variety of datasets in different application scenarios. Because the limitation of RF algorithm in measuring variables importance is that it does not automatically select the optimal number of variables that produce the best classification accuracy (Adam et al., 2009). Moreover, Genuer et al. (2015) reported some classification problems when the variable importance index is based on the Gini Purity index of RF. Previous studies also have demonstrated that the overall classification accuracy of the RF classifier decreases when the algorithm is trained in different study areas (Vetrivel et al., 2015). Later, Juel et al. (2015) tested the transferability of RF classification models and concluded that the resulting classification model was not transferable to new areas. Species composition in saltmarsh varies from one area to another due to different factors. Hence, the development of hybrid classification methods that integrate the RF classifier with explicitly specified models that convey the objects' semantics (Du et al., 2015) might be a solution to this problem.

There are some valuable investigations into the predictive power of a variety of variables in different classification scenarios. But the problem is the identified variables are not systematically organized into online features catalogues that can be easily reused by those interested in pursuing similar investigations in remote sensing community. In addition, Strobl et al. (2007b)argued that there is a bias in Random Forest variable measures in situations where potential predictors vary in their scale of measurement or their number of categories. For example,the increasing number of variables computed for the image objects generated through Object-Based Image Analysis (OBIA) classification(Belgiu and Drăguţ 2016) might, therefore, benefit from using the VI measurement to select optimum features.

However, the most important question that requires future investigation is the robustness of the RF classification model when the number of training samples is either increased or reduced (Chan and Paelinckx 2008). When noise is added to the features, the stability of the VI measure with iterative classifications (Millard and Richardson 2015), and the sensitivity of the *mtry* and *ntree* parameter to the feature space (i.e. to the number of variables) need further investigation. Recently, Mellor et al. (2015) reported that the correlation between trees increases as the number of variables for splitting the trees nodes increases. Therefore, feature selection and sample size still need further investigations to improve the stability of RF classification.

4.5 Concluding Remarks

Most of the findings show that there is empirical evidence to support the theoretical formulation and motivation behind both of the machine learning algorithms, SVM and Random Forest. However, the most important features of SVM is its ability to generalize well from a limited amount of training data. Compared to other non-parametric methods, SVMs can yield comparable accuracy using a much smaller training sample size. This is due to the ''support vector'' concept that develops based on a few data points to define the classifier's hyperplane. As the acquisition of ground truth for remote sensing data is generally an expensive process, SVM can be a good choice that works based on small sample size. In addition, SVMs do not assume a known statistical distribution of the data to be classified. This is very important for remotely sensed imagery that usually has unknown distributions. Therefore SVMs outperform parametric Maximum Likelihood Classification (MLC). Because MLC follow normal distribution and normality does not always give a correct assumption of the actual pixels distribution in each class to be classified.

On the other hand, RF classifier is less sensitive than SVM to the quality of training samples and to overfitting, due to a large number of decision trees produced by randomly selecting a subset of training samples and a subset of variables for splitting at each tree node. That's why, RF classifier has been shown to be suitable for classifying hyperspectral data, where the curse of dimensionality and highly correlated data pose major challenges to other available classifiers including SVM. In addition, RF has VI measurement that used to identify the most suitable seasons for identifying desired classes. Moreover, a large number of the investigations reviewed that the sensitivity of the RF classifier is based on two parameters (*ntree* and *mtry*) that are less than the number of parameters of SVM.

In a nutshell, it can be concluded that although SVM can deal with small sample size, less number of parameters settings and reliable variable selection method make RF easier to apply in image classification. In addition, sample proximity measurement available with the RF algorithm identifies outliers in training sample that is important when classifying objects in complex environments, i.e. wetland ecosystem.

Chapter 5

Support vector machine (SVM) classifier with small training samples for wetland saltmarsh environment

This chapter is based on

the article that has been submitted to the journal Geocarto International "*Support vector machine (SVM) classifier with small training samples for wetland saltmarsh environment".*

Abstract

Coastal saltmarsh is highly diverse, floristically and spatially heterogeneous, factors that make it challenging for remote sensing to map at species level. In addition, ground truth data collection for species level mapping is made challenging by limited access and hazardous conditions in some wetland ecosystems. Support Vector Machine (SVM) was selected for species level mapping for its unique behaviour with limited reference sample data. The accuracy of classification for coastal saltmarsh resulting from SVM is compared with that of the Maximum Likelihood Classifier (MLC). With a limited number of training sample, the overall accuracy (OA) for 8 classes was around 56.25% (kappa = 0.50) for MLC, 78.12 % (kappa=0.75) for SVM (radial basis function) and 78.90% (kappa=0.76) for SVM (polynomial). The results show SVM (either RBF or polynomial kernel) provide significantly (95% confidence interval) better results than MLC. There was no significant difference between the two maps produced from two different kernel functions at a 95% confidence interval. The relationship between kernel smoothness parameter and spectral separability are also investigated. When the polynomial kernel increased from 2 to 4, producer accuracy (%) increased from 81.25% to 87.50% and 53.22% to 66.67 % for Mangrove and Casuarina respectively. This accuracy is acceptable as only 15% of the required sample provided for 79% overall accuracy from SVM and is comparable to other previous studies where multispectral and airborne hyperspectral data have been used. However, in terms of data acquisition cost, reference data collection and processing cost, this study showed a promising approach for broad management application.

Keywords: *saltmarsh, classification, maximum likelihood classifier (MLC), support vector machine (SVM), Worldview 2, multispectral.*

5.1 Introduction

High spatial and spectral resolution remote sensing data with more advanced geospatial technology allows for the mapping of many changes in vegetation cover using species signature analysis. Some authors used airborne hyperspectral data such as PROBE -1, Compact Airborne Spectral Imager (CASI) imagery for mapping and monitoring salt marshes (e.g. Zomer et al., 2009; Belluco et al. 2006; Hunter and Power 2002; Thomson et al. 2003). However, aerial data acquisition is still considered relatively expensive (Hunter and Power 2002). Alternatively, high-resolution satellite imagery (HRSI) data and its recent advancement have the potential to significantly improve coastal and saltmarsh vegetation mapping. Due to the sub-meter spatial resolution and the advantage of satellite platform for repeated data acquisition with the minimal coast, Space Imagines' IKONOS and Digital Globe's Quickbird-2 have facilitated the routine change detection monitoring of both salt-marsh and terrestrial vegetation. For example, with high-spatial resolution QuickBird-2 satellite remote-sensing data, Wang (2007) mapped both terrestrial and submerged aquatic vegetation communities of the National Seashore Suffolk County, New York, using a combination of stratified and ISODATA classifier based on a previously classified map. The results achieved approximately 82% overall classification accuracy for terrestrial and 75% overall classification accuracy for submerged aquatic vegetation and provided an updated vegetation inventory and change analysis results.

One of the main purposes of any land cover classification is to achieve a high level of accuracy with a limited number of training samples to make mapping as practical and economical as possible. However, field data collection is not only time consuming and costly but potentially hazardous in the wetland environment. As a result, it is important to select a classifier that performs well with a limited number of ground reference samples. One attractive classifier for this application is a support vector machine (SVM) (Mathur and Foody, 2008). Literature shows that scientists have improved SVMs to successfully work with a limited quantity of training samples. Foody and Mathur (2004), for example, showed that limited training data (only 37 samples of the original 150 training samples) collected from SPOT high-resolution satellite imagery was enough to give the same accuracy for a two-crop classifier. In another study, Mantero et al. (2005) estimated probability density of thematic classes using SVM and confirmed the method effectiveness based on a limited number of ground reference samples. In addition, Bruzzone et al (2006) showed that modified SVM was effective in solving the ill-posed remote sensing classification problems associated with limited quantity and quality of training samples. As SVM is a supervised classification, it separates classes by identifying the support vectors from training samples. However, for a very large number of training samples, it is sometimes impossible for SVM to use all of them to determine support vectors (Koggalage & Halgamuge, 2004). In addition, a large training set size can limit the speed of SVM. As mentioned before, one of the major obstacles for wetland classification is accessibility to

collect ground samples, and classifier SVM might be an alternative to improve classification problem based on a limited number of ground truth data.

As SVMs are appealing due to their impressive ability to successfully handle a small training dataset by producing higher classification accuracy, it is proposed as a superior classifier in remote sensing than the traditional methods like Maximum Likelihood Classifier (MLC) (Montero et al., 2005), because, MLC is a probabilistic algorithm and less suited when data are non-normally distributed. The MLC classifier assumes that reflectance values of each class are normally distributed. This is a common violation in remote-sensing data, especially when classes or even subclasses of the main class contain different spectral features (Kavzoglou & Reis, 2008). To overcome this problem, non-parametric classifiers such as decision tree classifier (DTC), artificial neural networks (ANN), and Support Vector Machine (SVM) are gaining priority in recent remote sensing classification (Zhu and Blumberg 2002; Kavzoglu and Reis 2008; Otukei and Blaschke 2010). Among these, SVMs has been reported as the superior classifier in the work of Heinzel and Kotch (2012); Pal and Mather (2005); Huang et al., (2002) and Chen and Ho (2008). A comprehensive introduction of SVM can be found in the previous studies (Vapnik & Kotz, 1982; Zhu and Blumberg, 2002; Tso and Mather, 2009 and Vapnik, 2013) therefore it will not be further explained in this paper.

There is a close relationship between the choice of the kernel function and the performance of SVM (Tso & Mather, 2009). According to Huang et al., (2002), the function of a kernel has a major contribution in locating complex decision boundaries between classes. In any land cover classification, either wetland or terrestrial, one of the most important criteria is the distribution pattern of data. If the data is linear, a separating hyperplane may be used to separate the data into classes. However, it is common in remote sensing that boundaries between classes of interest are more likely to be non-linear and due to this linear kernel performed worse than nonlinear kernels (Huang et al., 2002). The solution is to use kernels in a higher dimensional space when solving the linear hyperplane. These are used to non-linearly map the input data to a high dimensional space. Therefore, the new mapping becomes linearly separable. As the kernel technique allows the support vector machine to form non-linear boundaries, different kernel functions (Cristianini & Taylor, 2000) have been introduced for SVM. Of these, polynomial and radial basis function (RBF) are two important functions used for remote sensing classification.

The distribution of classes on feature space is another important parameter that is related to the classification accuracy. This is because class probabilities are produced based on the distribution of classes in feature space. However, this distribution often has overlaps which reduce the accuracy of classification. To solve this problem, the concept of spectral separability is introduced. This concept indicates how well two classes are separated. This separability concept is classical in pattern recognition and independent of the coordinate system (Fukunaga, 1990; Fukunaga 2013) and has an effect on classification accuracy.

Based on these considerations, this study focuses on the potential of high-spatial-resolution satellite data for reliably classifying salt-marsh vegetation species with higher accuracy from limited field survey data. The main questions addressed in this study include: (1) how training size and parameter optimization influence SVM for species classification in a wetland? ; (2) how spectral separability index influences the kernel smoothness value and ultimately the classification accuracy?; (3) what mapping accuracy for individual saltmarsh species can be achieved? The results can then be used as baseline information for other related studies to develop a spatial data-based method of monitoring the condition of saltmarsh species in terms of degradation and rehabilitation.

5.2 Materials

The study area is dominated by Salt coach (*Sporobolus virginicus)*, and reedbeds of *Phragmites australis* (figure 5-1). Coastal swamp Oak (*Casuarina glauca*) and the grey mangrove tree (*Avicennia marina*) are two dominant tree species. In addition, there are some scattered saltmarsh species including *Saueda australis* and *Sarcocornia quinqueflora*. Due to their scattered distribution, those species were not classified as individual species in this study.

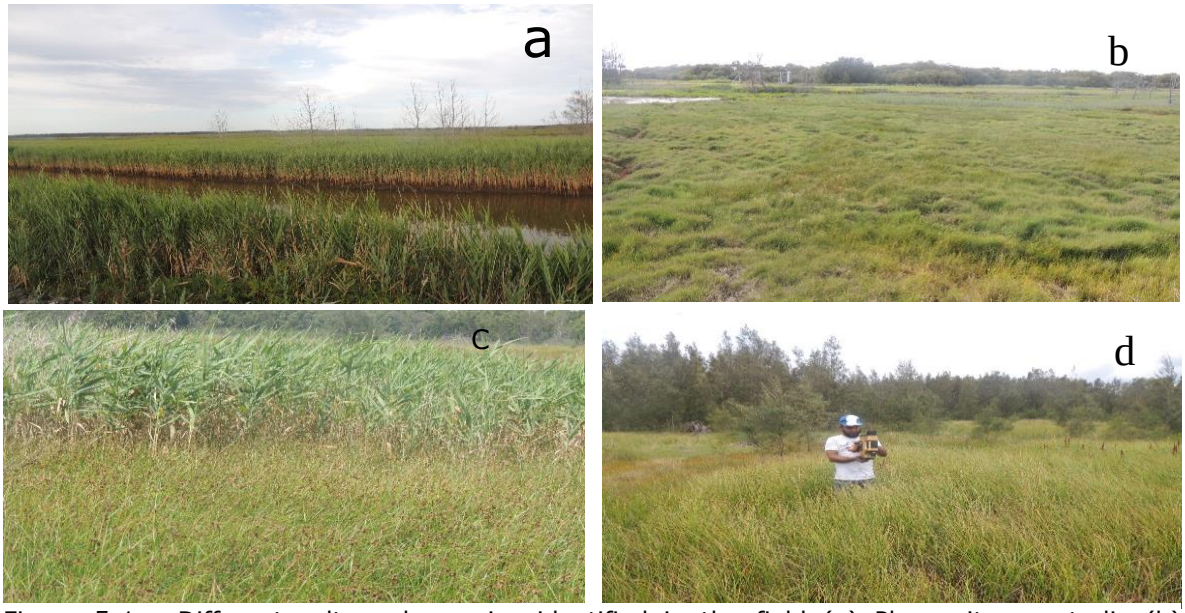

Figure 5-1: Different saltmarsh species identified in the field (a) Phragmites australis (b) Sporobolus virginicus (c) Mixed Phragmites and Schoeloplectus (d) Schoeloplectus sp mixed with tree Casuarina glauca (She-Oak).

5.2.1 Satellite and field data

Worldview-2 satellite imagery was the primary remote sensing data used in this analysis. It has 0.46 m pixel resolution in the panchromatic mode and 1.84 m resolution in the multispectral mode. The multispectral mode consists of eight broad bands in the coastal blue (400-450 nm), blue (450–510 nm),green (510-580 nm), yellow (585- 625 nm), red (630-690 nm), red edge (705 – 745 nm), NIR1 (770-895 nm) and NIR2 (860-1040 nm) parts of the electromagnetic spectrum. The Worldview-2 satellite data used here was captured on 5th May 2015. Two extensive fieldwork campaigns were conducted in two different seasons to check the variation in phenology of the target species if any. These are conducted from 10th to 14th June 2015, and from 7th to 12th December 2015.

5.2.2. Collection and processing of field data

The stratified sampling design was followed based on trees, saltmarsh and grass and others (water and wetland soil) (table 5-1). Within each stratum, reference pixels were randomly selected for calibration and validation purposes. Although homogeneity was a crucial issue for sampling size, however, each of the sample sites was more than 2 m x 2 m so that the data collected could be matched with the non-pan-sharpened pixels (2 m x 2 m) of Worldview-2. Sampling data included vegetation species class, percentage occurrence of each species within the selected plot and their global positioning system (GPS) locations. A total of 256 pixels were chosen from three different strata manually identified (Table1). Randomly collected ground reference sample from three different strata was divided into two groups, training and validation set. Hence, 50% pixels (128 Pixels) were selected for training and rest 50% (128 pixels) to calibrate the classifier algorithm (table 5-1).

Table 5-1: Different land cover pixels selected for training and validation dataset

Land cover	Specific Class	Calibration pixels (N= 128)	Validation Pixels (N= 128)
Trees	Mangrove	16	16
	Casuarina	16	16
	Casuarina dieback	16	16
Saltmarsh and Grass	*Phragmites australis*	16	16
	Sporobolus virginicus	16	16
	Perennial Grass	16	16
Others	Water	16	16
	Wetland Soil	16	16

5.2.3 Training data set

Selection of training data is one of the major factors determining the extent to which the classification rules can be generalized to unseen samples over the study site (Paola and Schowengerdt 1995). This selection of training pixels could be me more important for obtaining accurate classifications than the selection of classification algorithms (Hixson et al. 1980). To assess the impact of training data size on classification accuracy, an equal number of samples for each class were trained and validated with all bands of Worldview-2. As mentioned, 50% samples (128 pixels) were separated for training purposes where each class has an equal number of pixels (16 pixels for each class). On each trial, validation sample number was equal to the training sample number but using independent sets. There were 9 trials and the sample size was increased from 8 to 16 with a step size of 1. This randomization technique was used to minimize the effect of spatial autocorrelation (Campbell 1981) and was the best approach to check the effect of training sample on SVM with limited sample number.

5.2.4 SVM Kernel size

The kernel parameter of SVM has a significant effect on the decision boundary of two classes. The width parameter of the Gaussian Kernel and the degree of the polynomial kernel control the flexibility of the resulting classifier (Ben-hur & Weston, 2010). The lowest degree polynomial is the linear kernel and is not sufficient if there is a non-linear relationship between the features to be classified. In ENVI, the minimum value is 1 (default), and the maximum value is 6 for the degree of the polynomial kernel. Increasing this parameter more accurately delineates the boundary between classes (Harris Geospatial Solution, 2015). A value of 1 represents a first-degree polynomial function, which is essentially a straight line between two classes. This value works well if there are two distinctive classes in an image. However, in most cases, land cover classification deals with imagery that has a high degree of variation and mixed pixels. Thereby increasing the polynomial value causes the algorithm to more accurately follow the contours between classes (Harris Geospatial Solution, 2015).
Besides choosing a kernel type, it is also essential to specify the bias in kernel function. In ENVI the default bias value is 1.00. Similarly, it is also important to set the gamma parameter used in the kernel function. This value is a floating value greater than 0.01. The penalty parameter is a floating-point value greater than 0.01. In ENVI the default penalty value is 100.0. This parameter allows a certain degree of misclassification, which is particularly important for non-separable training sets. This parameter controls the magnitude the penalty of training samples that lie on the wrong side of the hyperplane parameter. The more the user increases the value, the easier it is easy to control if making changes to other parameters (Harris Geospatial Solution, 2015). Increasing this value also

increases the cost of misclassifying points and creates a more accurate model that may not generalize well (Tso & Mathur, 2009). In SVM, the best kernel functions are data dependent (Beh-Hur and Weston, 2010). Therefore it is imperative to try different kernel functions with an associated parameter to select the best one. In this study, different kernel functions were tested to select the optimal one for wetland mapping.

5.2.5 Class separability

There are several methods to identify class separability. For example, Jia & Richard (1999) mentioned the divergence and Jeffries-Matusita (JM) distance to measure class separability. In this research Jeffries-Matusita(J-M) distance was used to measure the separability between the 8 classes chosen. The value of this distances varies from 0 to 2.0 where 0 indicates classes are same and 2.00 indicates they are very well separable.

5.2.6 Evaluation of classification and mapping accuracy (MA)

Accuracy assessments were conducted using the confusion error matrix, overall map accuracy and kappa values. For individual classes, producer and user accuracy were computed based on the dominant class in each reference plot (Story and Congalton 1986; Congalton 1991; Richards 1996; Stehman 1997). In addition, MA [Equation 5.1] (Congalton and Green 1999) was computed based on the following formula:

$$MA\% = \frac{Pixels_{Correctly\ classified}}{Pixels_{Correctly\ classified}+Pixels_{Omission}+Pixels_{Commission}} \times 100 \qquad (5.1)$$

The Z- statistics (Congalton and Mead 1986) was used to evaluate the significance of the difference between the resulting classifications obtained by two methods: SVM and MLC. For example, the difference between the two classifier or two individuals in two different classifiers is considered to be significant at 95% confidence level if the absolute value of the Z-statistics exceeded 1.96. The Z- Statistics [Equation 5.2] is calculated as below:

$$Z_{ab} = \frac{|K_a - K_b|}{\sqrt{Var_{K_a}+Var_{K_b}}} \qquad (5.2)$$

Where Z_{ab} is the Z-statistic for comparison of classification a and b; K_a and K_b are the kappa coefficients of classifications a and b; and var(K_a) and var(K_b) are the asymptotic variances of K_a and K_b, respectively. All statistical computing was performed in open source R software. (R core team, 2014).

5.3 Results

5.3.1 Data distribution pattern

The box-whisker plot in figure 5-2 indicates the spectral variability and data distribution pattern among and within 8 classes. Only annual grass (GR) is clearly separable in green, red edge and NIR bands. Reflectance properties of water (WA) in the last 3 bands (Red Edge, NIR1 and NIR 2) and reflectance of wetland soil (WS) in band 8 are significantly different from other classes. Within two tree species (Casuarina=CA and Mangrove=MA) the separation is not so clear, there are considerable spectral overlaps in different bands. Two saltmarsh species (Phragmites= PH, and Sporobolus = SP) are more or less overlapped in different bands. The saltmarsh species also show band-specific within-species variance. For example, the variance of Phragmites (PH) in the coastal band is quite large and in NIR2 is relatively small. Based on the overall box-whisker plot, it is clear that there are no outliers for any species in any band. Moreover, the median is in the centre of each box indicating that data are normally distributed.

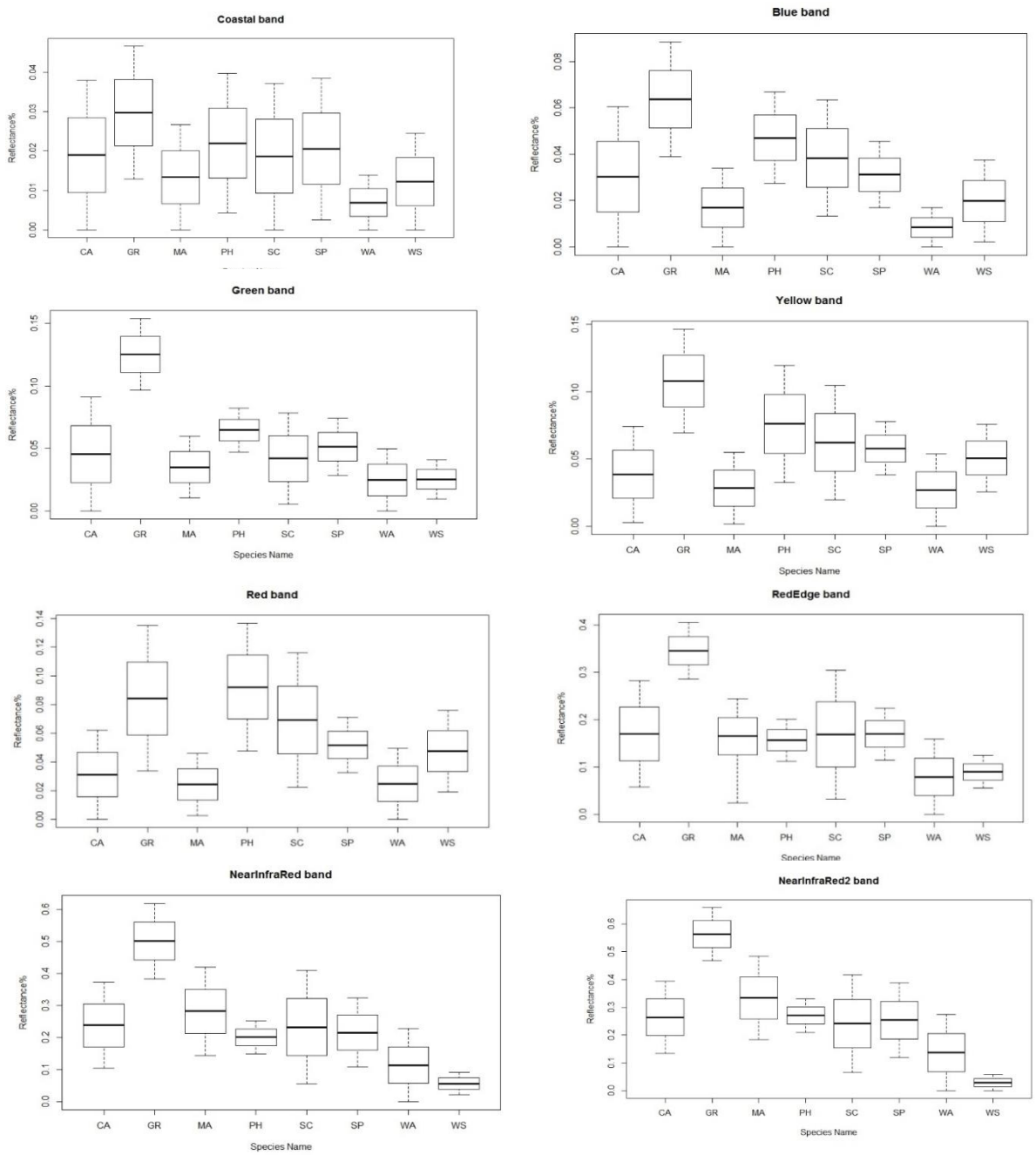

Figure 5-2: Box-whisker plots of median reflectance values of the 8 Worldview 2 bands for different wetland species derived from the reflectance pixels of the study area

5.3.2 Maximum Likelihood Classification (MLC)

The confusion matrix (table 5-2) of the MLC classifier with 16 training samples from each class indicates that there is a high incidence of misclassification of mangrove and Casuarina sp. This confusion mostly comes from the similarity of the spectral properties of these tree species (Figure 5-2). Another source of classification error come from the similarity of the spectral profile of Sporobolus virginicus with *Phragmites australis*.

Table 5-2: Confusion matrix of MLC classifier

Class	MA	CA	WA	PH	SC	SP	GR	WS	Total	EC	UA
MA	**9**	5	1	0	4	1	2	1	23	60.87	39.13
CA	3	**8**	0	0	0	1	0	0	12	33.33	66.67
WA	0	2	**11**	0	0	0	0	0	13	15.38	84.62
PH	0	0	0	**7**	1	0	0	0	8	12.50	87.50
SC	4	1	1	8	**10**	6	8	1	39	74.36	25.64
SP	0	0	0	1	1	**8**	0	1	11	27.27	72.73
GR	0	0	0	0	0	0	**6**	0	6	0.00	100.00
WS	0	0	3	0	0	0	0	**13**	16	18.75	81.25
Total	16	16	16	16	16	16	16	16	**128**		
EO	43.75	50.00	31.25	56.25	37.50	50.00	62.50	18.75	OA=56.25%		
PA	56.25	50.00	68.75	43.75	62.50	50.00	37.50	81.25	Kappa =0.50		

Class key: MA= Mangrove, CA = Casuarina, WA= Water, PH = Phragmites australis, SC = Casuarina dieback, SP= Sporobolus virginicus, GR = Grass, WS = Wetland Soil, EO = Error of Omission, EC= Error of Commission, UA= User Accuracy, PA= Producer Accuracy, OA = Overall Accuracy.

5.3.3 Support Vector Machine (SVM)

Determination of parameter

As described earlier, both RBF and polynomial kernels of SVM were tested. To determine the parameter for Radial Basis Function (RBF) in the SVM classifier, we did different trials of gamma in kernel function ranging from ENVI default value, 0.125 to 7.00. Based on this trial it was found that there was no change of OA when gamma varied from the default value to 7.0. Therefore the ENVI default gamma kernel value 0.125 has been selected for the RBF SVM classifier. In respect of the penalty parameter, the ENVI default value is 100. Other parameters were tested ranging from value 5 to 200 to check the effect of penalty parameter in classification accuracy and kappa followed by a McNamara test. It was proved

that OA increased from 72.65% (Kappa = 0.68) to 78.12 % (Kappa =0.75) when the penalty parameter was raised to value 130. However, it remained unchanged up to value 140 and reduced once again when it was set as 150 or more. These differences in classification accuracy were significant at the 0.05 probability level. As a result, penalty value 130 with a game value 0.125 was treated as the best performance for RBF classification.

For the polynomial, varying order from value 2 to 4 resulted in an increase in overall classification accuracy from 77.34% (kappa = 0. 74) to 78.90 % (kappa= 0.75) but accuracy reduced once again when order values of 5 and 6 were used (OA 75.78 % with kappa 0.75). However, based on the McNamara test, this difference was not significant at the 0.05 probability level. As polynomial order 4 was given the highest OA, this value was selected for the remaining trials of other parameters of the polynomial SVM classifier. Due to the change of penalty parameter from value 5 to 100, the OA increased from 71.87 (kappa = 0.67) to 78.90% (kappa = 0.75). But it reduced once again when penalty parameter set at more than 100. In this study, penalty parameters from 5 to 5000 were tested. However, based on the McNamara test the classification accuracy at a penalty value 5000 was not significantly different from the accuracies for other penalty values. Therefore, the highest OA 78.90% with kappa 0.75 obtained from polynomial kernel 4 and a penalty value 100 has been used for further comparison and mapping accuracy calculation.

Table 5-3: Confusion matrix of SVM (RBF) classifier

Class	MA	CA	WA	PH	SC	SP	GR	WS	Total	EC	UA
MA	**14**	4	0	0	2	0	0	0	20	30.00	70.00
CA	2	**9**	0	0	0	0	0	0	11	18.18	81.82
WA	0	3	**14**	0	1	0	0	1	19	26.32	73.68
PH	0	0	0	**13**	2	1	0	0	16	18.75	81.25
SC	0	0	1	1	**6**	1	1	0	10	40.00	60.00
SP	0	0	0	2	5	**14**	0	0	21	33.33	66.67
GR	0	0	0	0	0	0	**15**	0	15	0.00	100.00
WS	0	0	1	0	0	0	0	**15**	16	6.25	93.75
Total	16	16	16	16	16	16	16	16	**128**		
EO	12.50	43.75	12.50	18.75	62.50	12.50	6.25	6.25	OA=78.12%		
PA	87.50	56.25	87.50	81.25	37.50	87.50	93.75	93.75	Kappa =0.75		

Class key: MA= Mangrove, CA = Casuarina, WA= Water, PH = Phragmites australis, SC = Casuarina dieback, SP= Sporobolus virginicus, GR = Grass, WS = Wetland Soil, EO = Error of Omission, EC= Error of Commission, UA= User Accuracy, PA= Producer Accuracy, OA = Overall Accuracy.

Table 5-4: Confusion matrix of SVM (Polynomial) classifier

Class	MA	CA	WA	PH	SC	SP	GR	WS	Total	EC	UA
MA	**14**	3	0	0	2	0	0	0	19	26.32	73.68
CA	2	**10**	0	0	0	0	0	0	12	16.67	83.33
WA	0	3	**14**	0	1	0	0	1	19	26.32	73.68
PH	0	0	0	**13**	2	1	0	0	16	18.75	81.25
SC	0	0	1	1	**6**	1	1	0	10	40.00	60.00
SP	0	0	0	2	5	**14**	0	0	21	33.33	66.67
GR	0	0	0	0	0	0	**15**	0	15	0.00	100.00
WS	0	0	1	0	0	0	0	**15**	16	6.25	93.75
Total	16	16	16	16	16	16	16	16	**128**		
EO	12.50	37.50	12.50	18.75	62.50	12.50	6.25	6.25	OA=78.90%		
PA	87.50	62.50	87.50	81.25	37.50	87.50	93.75	93.75	Kappa =0.76		

Class key: MA= Mangrove, CA = Casuarina, WA= Water, PH = Phragmites australis, SC = Casuarina dieback, SP= Sporobolus virginicus, GR = Grass, WS = Wetland Soil, EO = Error of Omission, EC= Error of Commission, UA= User Accuracy, PA= Producer Accuracy, OA = Overall Accuracy.

From figure 3 it is very clear how the Phragmites australis patch has been misclassified by MLC and further pixels have been purified by SVM. Similarly, Table 5-2, 5- 3 and 5-4 prove that classification improved by application of SVM for all eight classes. From Table 3 and 4, it is true that accuracy slightly improved by application of the polynomial kernel, however, producer and user accuracy for all classes remain unchanged except Mangrove and Casuarina sp. The reason is that the smoothness parameter of polynomial kernel mainly works on the poorly separable classes, as discussed in next section.

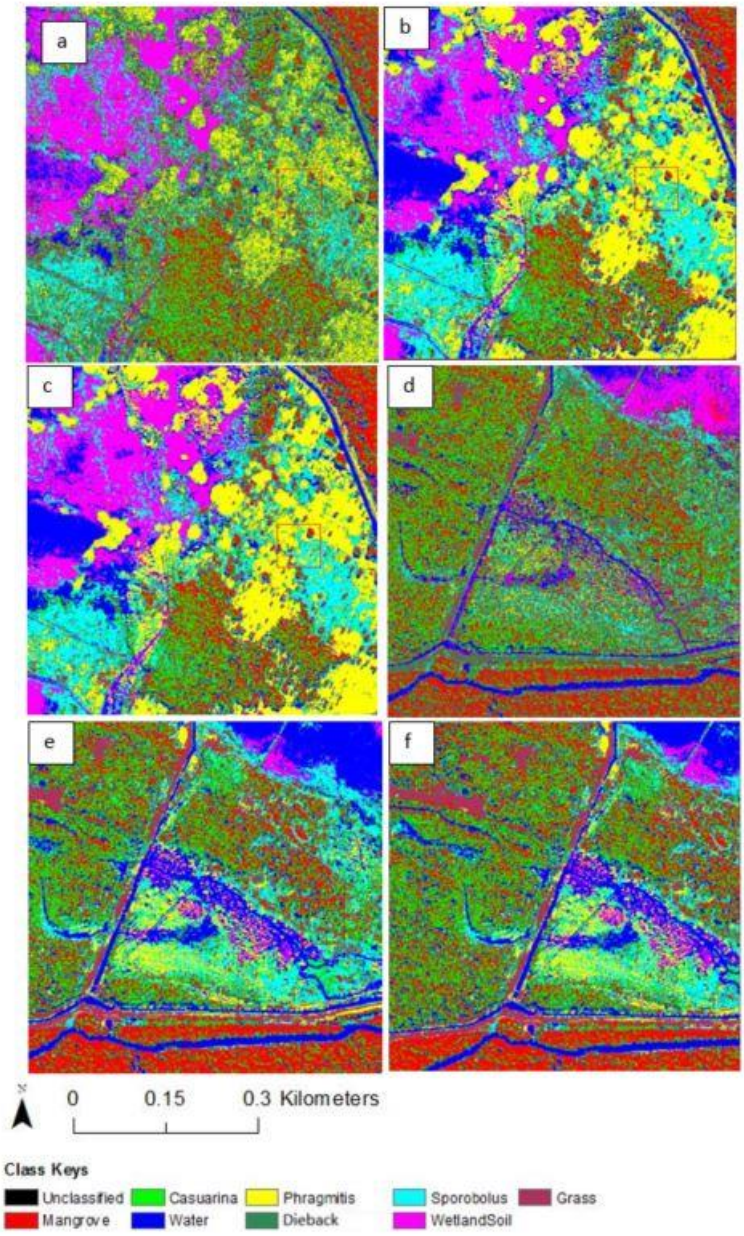

Figure 5-3: Classification results highlighting Phragmites australis, Sporobolus virginicus and Casuarina sp, (a) MLC (b) SVM-Polynomial (c) SVM-RBF; Classification highlighting Mangrove, Water Channels and Grass dominant area (d) MLC (e) SVM-polynomial (f) SVM-RBF

5.3.4 Relation between class separability and kernel smoothness parameter

Based on different training pixels, it was found that the value of separability index depends on the number and choice of pixels (table 5-5). For example, the J-M separability index for Mangrove and Casuarina sp, varied from 1.48 to 1.99 based on different pixels number. For this reason, the relation between class separability (J-M) and kernel function was tested in terms of polynomial degree.

Table 5-5: Class separability and Kernel function

Separability index (J-M)		Training Set		Kernel function
Mangrove	Casuarina	Trial Number	Pixel	Polynomial Degree
	1.99	3	10	2 to 6
	1.93	5	12	2 to 6
	1.91	6	13	2 to 6
	1.77	7	15	2 to 6
	1.48	8	16	2 to 6

Different degree of the polynomial kernel was tested followed by contingency table classification accuracy (table 5-6). It was found that the order of polynomial degree worked for poorly separable classes (for example, Casuarina and Mangrove) and improved classification accuracy. According to expectation, for well separable class (higher J-Mvalues) classification accuracy improved and sensitivity to degree function reduced. Conversely, for lower class separability (Mangrove-Casuarina; table 5-6), the role of smoothness parameter of degree becomes higher and accuracy also changed. From the confusion matrix table (table 5-6), it is also evident that found that higher degree value works to separate poorly separable class and thereby improve classification accuracy. When 16 a pixel size training set was tested against varying polynomial degree, PA % and UA % both improved with the change of degree. This argument is also supported by Goumehei (2010) who found the similar relationship of JM separability distance with kernel smoothness parameter.

Table 5-6: Relation between kernel function and classification accuracy

User accuracy (%)		Producer accuracy (%)		Overall accuracy	Polynomial degree	Pixel size
Mangrove	Casuarina	Mangrove	Casuarina			
72.22	76.92	81.25	53.22	77.34	2	16
73.68	83.33	87.5	66.67	78.12	3	16
73.68	83.33	87.5	66.67	78.90	4	16
66.67	80	87.5	66.67	75.78	5	16
76.47	81.82	81.25	60	76.56	6	16

5.3.5 Mapping Accuracy Assessment

Table 5-7: Mapping Accuracy (%) of different wetland classes

Class	MLC	SVM(RBF)	SVM(Polynomial)
Mangrove	30.00	63.63	66.66
Casuarina	40.00	50.00	55.55
Water	61.11	66.66	66.66
Phragmitis australis	41.11	68.42	68.42
Casuarina dieback	22.22	30.00	30.00
Sporobolus virginicus	42.10	60.86	60.86
Grass	37.5	93.75	93.75
WetlandSoil	68.42	88.23	88.23

From Table 5-7, it is annual grass, achieved higher MA (93.75%) followed by wetland soil (88.23%) from the SVM classifier. Except for Mangrove and Casuarina sp. all other six classes remain the same in both kernel type of SVM (table 5-7). The reason is very clear from the separability value and smoothness parameter relationship that is discussed in the previous section.

5.3.6 Evaluation of performance

Table 5-8 shows the results of kappa analysis over three classification results. It is very clear that the overall accuracy and kappa coefficient by SVM (either RBF or Polynomial) are significantly (95% confidence interval) higher than that of MLC. Because in both cases Z –statistics are higher than 1.96 (table 5-9). Although SVM (Polynomial) provides slightly higher accuracy than SVM (RBF), however, their differences are not significant (95% confidence interval) since Z-statistics are smaller than 1.96 (table 5-9).

Table 5-8: Kappa analysis

	MLC	SVM(RBF)	SVM(Polynmial)
Overall accuracy	0.5625	0.7812	0.7890
Kappa coefficient	0.5011	0.7512	0.7589
Variance	0.001923	0.001335	0.001300

Table 5-9: Z-statistics among classification results

Pair comparison	Z-statistics
MLC vs SVM (RBF)	3.8325
MLC Vs SVM (Polynomial)	3.9908
SVM (RBF) Vs SVM (Polynomial)	0.0288

5.4 Discussion

The literature often suggests that the size of the training set required is a function of spectral wavebands used and generally a sample comprising at least 30 times the number of spectral bands used is required in the analysis process (Mather, 2004; Mathur and Foody 2008). Moreover, it is proved that training set size has a positive relationship with the classification accuracy (Zhuang et al., 1994, Arora and Foody 1997, Pal and Mathur 2003, Foody and Mathur 2004a). That is why conventional classification scheme (like MLC) require a large training set spread all over the study area. In this perspective, our Worldview 2 data requires at least 240 pixels for each class to run a conventional MLC algorithm. Whereas, only 15% (16 pixels for each class) of the required training sample (240 pixels) provided up to 79% accuracy. This accuracy is acceptable not only based on cost-benefit analysis but the feasibility test as well. For wetland in many settings, it was difficult and expensive to collect at least 240 pixels of training data for each class to test MLC performance. The findings of the current study are supported by Mathur and Foody (2008): when they reduced training size from 450 pixels to 130 pixels and classification accuracy reduced only 1.34% (statistically not significant). Although SVM provided 1.34% less accuracy (90.66%) compare to MLC (92.00%), Mathur and Foody (2008) concluded in favour of the superiority of SVM for small training size and cost of analysis.

The study differs from Carle et al., (2014) who suggested MLC performed better than SVM if data are normally distributed. Although our data were normally distributed, still SVM performed better than MLC. The main reasons are our high spatial resolution data allowing a small training set size. Jin et al., (2005) demonstrated that the difficulty in MLC dealing with texture features of high-resolution images, especially with a small training set size situations. In addition, MLC suffers from the Hughes effect or the curse of dimension (Hughes 1968, Pal and Mather 2003). This means that with a fixed number of training samples performance declines with the increasing dimensions due to the decreased reliability of estimates of statistical parameters needed to calculate the probability (Oommen et al. 2008). Our SVM findings challenge previous findings where highest classification accuracy has been achieved with the SVM classifier using a Radial Basis

Function (RBF) kernel. For example, Pal and Maher (2005) found optimum classifier performance both for mapping agricultural crops from Landsat ETM+ data and for mapping broad land-cover classes from hyperspectral data with RBF kernel gamma value 2, and a penalty value of 5000. Similarly, Kavzoglu and Colkesen (2009) found that the RBF kernel with a gamma value of 3 and penalty value of 250 was optimal for superior accuracy for broad land-cover classes using Landsat ETM+ and Terra ASTER imagery. However, their findings also support the argument that it is essential to optimize each kernel based on classification task and imagery type. So kernel determination depends on image type and classification task.

Our findings are similar to Dixon and Candade (2008) who obtained optimal results using the polynomial kernel with an order of 3 and penalty value of 1000 from Landsat TM 5 data. Our finding also supported by the findings of Huang et al (2002) who applied this algorithm to two different satellite data sources. So it may be summarized that optimal kernel choice is also highly variable among remote-sensing applications. This results also concur with other coastal vegetation studies where high-resolution data has been used. For example, Collins and Planes (2011) achieved more than 0.90 kappa coefficient with SVM from Worldview 2 data. It is true that our accuracy is lower than Collins and Planes (2011), however, this lower accuracy is mainly due to our exclusive focus on individual species where spectral signatures overlapped. When Collins and Planes (2011) showed 28 pairwise comparisons in their land cover classification, only 1 showed a separability index inferior to 1.99. In our dataset, out of the 28 pairwise comparisons possible, 13 pairs displayed an inferior index to 1.99 (2.00 is a perfect discrimination) [See Appendix 1] and acted as an obstacle to improve accuracy after a certain degree. Dixon and Candade (2008) and Immitzer, Atzberger, and Koukal (2012) similarly found user's accuracies ranging from 57% to 100% and producer accuracies ranging from 33% to 100% for individual species using Landsat 5 TM and Worldview 2 respectively. In addition, our accuracy makes a cost-effective effort with the limited sample and less expensive imagery compare to the accuracies achieved for species-specific classification using hyperspectral imagery (Filippi and Jensen 2006; Belluco et al. 2006).

5.5 Limitations and further research scope

It is imperative to use a grid search method for parameter optimization that provides the best value for all parameters using the training dataset with a cross-validation approach. In this experiment, we used a random value for parameters in ENVI software. Feature selection for SVM although introduced in chapter 4, however, have not used yet. SVM has a large number of parameters that need to be optimized. So its performance needs to be compared with another ensemble classifier that has less number of parameters. Lastly, the proposed training sample size is restricted to 16 pixels for each class. Therefore, it is

essential to test the impact of larger training sample size that will prove the robustness of SVM's performance.

5.6 Conclusion

The wetland area selected for this research has a unique feature in terms of the construction of levies for management purposes, prior history of disconnection of water channels from the sea, and a subsequent restoration. This has resulted in degradation of saltmarsh habitat, complex patterns of species invasion following hydrological alteration, and subsequent dieback and restoration of saltmarsh of saltmarsh habitat in some locations. The mosaic of saltmarsh and other coastal wetland vegetation patterns represented at Tomago provides a challenging test vegetation classification and mapping. We were particularly interested in testing the effect of limiting reference collection given the impediment this is placed on global saltmarsh mapping to date. Our challenge was to find a good classifier that works well with limited training data for saltmarsh wetland classification. To test the strength of conventional MLC performance, the required training size (240 pixels for each class) was unavailable in this study and acknowledged limitation. However, when 15% of the required training sample provided up to 79% accuracy from SVM was achieved, a result is acceptable in respect of cost and feasibility analysis for wetland environment. The SVM (either RBF or polynomial kernel) clearly outperform the traditional MLC especially when the training dataset has a small size of 16 samples. In terms of less expensive data cost and processing time, these SVM accuracies are competent and rival those findings using similar multispectral and airborne Hyperspectral sensors. Relative to the use of conventional MLC classifier, SVM provides 21.87% more accurate result with small sample size and provides a significant cost advantage. From the result of spectral separability, it was also found that higher degree kernel value works to separate poorly separable class. For example, when polynomial kernel increased from 2 to 4, producer accuracy increased from 81.25% to 87.50% and 53.22% to 66.67 % for Mangrove and Casuarina respectively. Although support vector machine requires some effort and time for kernel and parameter optimization for specific remote sensing application, it is a good choice to provide a satisfactory result when reference data collection is a big issue.

Chapter 6

An evaluation of equal training sample allocation for a saltmarsh environment using Random Forest (RF) and Support Vector Machine (SVM) Classification algorithms

This chapter is based on

the article that has been submitted to the Special issue on "Marine Protected Areas: Science, Policy & Management" of International Journal of Estuarine, Coastal and Shelf Science.

Abstract

Ground truth data collection for species level mapping is made challenging by limited access and hazardous conditions in some wetland ecosystems. Instead of area based proportional sample allocation, an equal number of samples allocation strategies have been proposed for 8 classes within a saltmarshes community. Random Forest (RF) and Support Vector Machine (SVM), two machine learning algorithms were selected for species level mapping due to their unique behaviour with multi-modal data distribution and limited reference sample size. Only RF showed significant changes in overall accuracy when the sample size was reduced from 100% level to 33% : accuracy dropped from 79% to 72 %. Conversely, there were no significant changes in the accuracy of SVM, when samples size equally dropped from 100% to 33%. Based on equal sample allocation strategies, only 15% of the required samples provided for 75% of the overall accuracy of SVM. However, there was no significant difference among all the possible combinations of three experiments (100 %, 66% and 33% of the original samples) which originated from two classifiers based on an equal sample distribution scheme. The importance of each spectral bands was evaluated through Classification and Regression Training (CARET) and Random Forest (RF) packages in R environment. Learning Vector Quantization (LVQ) and Recursive Feature Elimination (RFE), both feature selection methods selected six important features. No significant differences in classification accuracy were observed when these six important features were selected instead of all available features. This accuracy is acceptable and is comparable to other previous studies where multispectral and airborne hyperspectral data have been used. However, in terms of data acquisition cost, reference data collection and processing costs, this study showed a promising approach toward broad management application for the saltmarshes community that is ecologically endangered, and demand a special attention.

Keywords: *saltmarsh, wetland, maximum likelihood classifier (MLC), machine learning, Worldview 2, multispectral.*

6.1 Introduction

Thematic mapping of any land cover from remotely sensed data is commonly based on image classification methods. Lu and Weng (2007) provide a review of commonly used classification methods applied to remotely sensed data and these methods can be divided into common and advanced image classification (Tso and Mather 1999). For example, the Maximum Likelihood Classifier (MLC) is a commonly used supervised classifier whereas ISODATA and K-Means are unsupervised methods (Richards and Jia 1999; Srivastava et al. 2012). Advanced classification algorithms include support vector machines (SVM), random forest (RF), artificial neural networks (ANN), and Decision tree classifier (DTC)(Adam et al. 2014; Adam et al. 2010). The performance of these classifiers varies with the quality of the remote sensing data either passive sensor (multispectral or hyperspectral data) or active sensor (e.g. LiDAR) data. One of the major problems relating to the supervised classification lies in the definition of a proper training set size for an accurate learning of classifiers(Chi et al. 2008).

Sample size balance maintaining is still a controversial topic in image classification. It could be argued that classes with multimodal frequency distributions, e.g., agricultural land with different crop types and growing cycles, should have a greater number of samples in order for classes to be accurately represented in the classifier than a spectrally and temporally well-defined class i.e. waterbody (Colditz 2015). Accordingly, classification trees may sometimes suffer from a problem of unbalanced sample sizes problem. This means in the standard form of classification trees, the class with the highest number of samples determines the class label (Colditz 2015).

To analyze the sample size problem for SVM and RF, we synthesized two review works of SVM and RF done by Mountrakis et al. (2011) and Belgiu and Drăguţ (2016) respectively. Based on the review of RF, it can be concluded that there is still a research paucity / gap on sample size and mislabelling of data. However, Mellor et al. (2015) found that RF classification was relatively insensitive to mislabelled training data and that imbalanced training data can be introduced to reduce the errors in those classes that pose the greatest challenges to classifications. Conversely, Millard and Richardson (2015) and Dalponte et al. (2013) found that the RF classifier fails to cope with imbalanced training data and tends to favour the most representative classes. In another study, Jin et al. (2014) concluded that the proportionally allocated training sample design reduces the commission error of the under-represented classes and that the equally allocated training sample schema reduces the omission error of the under-represented classes. In these circumstances, the impact of sampling design on RF classification results seem to be contradictory. Hence subsequent studies are required to analyse the sensitivity of the RF classifier to training samples allocation when using this classifier for remote sensing data classification.

As SVM for remote sensing classification has reviewed (Mountrakis et al. 2011) several years ago, we additionally synthesized most of the recent works in which SVM was used to classify land cover or wetland vegetation. Accordingly, we found that Carle et al. (2014) observed MLC performed better in comparison to SVM when wetland vegetation data are normally distributed. Recently,Fang et al. (2016) also found a better result from a decision tree classifier (DTC) than SVM when they classified wetland from multisource data. Zhang et al. (2015) used multi-parameter optimization (MP-SVM) that is a modified form of single parameter optimization (SOP-SVM) and achieved a better result. It was also revealed that single optimized parameter (SOP) and a novel multi-parameter (MP) both are sensitive to landscape to be classified. As a result, evaluation of SVM parameter performance based on different sample sizes is still a current research gap that can be assessed in a saltmarsh environment where single to multi-modal distributions are prevalent.

Since 1996 RF and SVM has been widely explored in Land Use Land Cover (LULC) classification utilising commonly used multi-spectral imagery. Recently Adam et al. (2014) applied these two algorithms on multispectral RapidEye data for a heterogenous landscape classification. Although RapidEye has five multispectral bands including Red-edge and NIR 1, there is a gap in knowledge of the performance of these algorithms on high-resolution Worldview-2 imagery. Worldview-2 (WV-2) can be spectrally differentiated from RapidEye as the former has three additional bands, called Coastal blue (Band1), Yellow (Band 4) and Near Infrared Band II (Band 8). Moreover, there is no mentionable research where these two algorithms were compared to examine the proportional sample reduction on user and producer's accuracy for a saltmarsh environment. Consequently, this study sought to compare the performance of the relatively new advanced machine-learning RF and SVM classifiers on the new-generation WV-2 imagery in a degraded saltmarsh wetland to determine the optimum sample size for classification and mapping purposes. We focus on the following research questions (1) How do machine learning algorithms RF and SVM respond to the proportional reduction (100 % to 33%) of sample size? (2) Which machine learning approach provides reasonable accuracy with limited sample number? (3) Do the selected features of WV-2 change classification accuracy significantly compared to using all bands of WV-2?

6.2 Materials

6.2.1 Collection and processing of field data

There are three major strata in the study area i.e. trees, saltmarsh and wetland. Within each stratum, reference pixels were randomly selected for calibration and validation

purposes. As pixel number was the main focus of this study, we did not collect sample classes based on their represented sizes within the study area but we allocated equal pixels numbers to all classes within each stratum. Two extensive fieldwork campaigns were conducted in two different seasons (winter and summer) to check the variation in phenology of the target species if any. These were conducted from the 10th to 14th of June 2015, and from the 7th to 12th of December 2015.

A total of 1536 pixels were chosen from eight different manually identified classes (Table 6-1). Randomly collected ground reference samples from these eight different classes were divided into two sets: training and validation sets. Hence, 50% pixels (768 pixels) were selected for training and the rest 50% (768 pixels) was used to calibrate the classifier algorithm. The validation set was performed without including any samples inside the eight classes of the training subset. Therefore, the bias in the accuracy evaluation due to the use of the same data sets of pixels for both calibration and validation was avoided. Ground-based field samples were then collected from each class by traversing the study area where access was possible. Additional reference data for hard-to-reach locations were obtained by visual interpretation of Near map, NSW "six map" spatial database website and Office of Environment and Heritage (OEH) database.

Table 6-1: List of three different experiment sets based on the pixel number

Experiment	Number of variables	Calibration	Validation	Total pixels for calibration and validation
1	8 spectral bands	Pixel No. =16	Pixels No. =16	16*8=128
2	8 Spectral bands	Pixels No. = 32	Pixels No. = 32	32*8=256
3	8 Spectral bands	Pixels No. = 48	Pixels No. = 48	48*8=384
Total				786 pixels

6.2.2 Processing of satellite data

Worldview-2 satellite imagery was the primary remote sensing data used in this analysis. It has a 0.46 m pixel resolution in the panchromatic mode and 1.84 m resolution in the multispectral mode. The multispectral mode consists of eight broad bands (Table 6-2). The Worldview-2 satellite data used here was captured on the 5th May 2015. The pixels value (Digital Number) were converted to 'at sensors' radiance and then we performed an atmospheric correction with the ENVI module (ENVI Classic) FLAASH to retrieve reflectance data.

Table 6-2: Summary of the variables derived from the Worldview 2 data used

Variables	Wavelength (nm)
Band 1 (Coastal bands)	400- 450 nm
Band 2 (Blue)	450 – 510 nm
Band 3 (Green)	510 – 580 nm
Band 4 (Yellow)	585 – 625 nm
Band 5 (Red)	630 – 690 nm
Band 6 (Red Edge)	705 – 745 nm
Band 7 (Near Infrared 1)	770 – 895 nm
Band 8 (Near Infrared 2)	860 – 1040 nm

6.2.3 Random Forest (RF)

RF is an ensemble learning technique that uses a set of CARTs to make a prediction (Breiman 2001). This algorithm creates a subset of training samples through replacement (a bagging approach).Therefore, the same sample can be selected several times, while other samples may not be selected at all. Usually, two-thirds of the samples (referred to as in-bag samples) are used to train the trees (*ntrees*). The remaining one third (referred to as out-of-the-bag samples) is used in an internal cross-validation technique for estimating how well the resulting RF model performs (Breiman 2001). This means that the OOB sample (one-third of the total sample) is used to estimate the misclassification error (OOB error) and variable importance. At each node, a given number of input variables (*mtry*) are randomly chosen from a random subset of the features and the best split is calculated by utilizing only this subset of features. RF is now a widely used algorithm for remote sensing image classification (Ozesmi and Bauer 2002). Its ability to handle high dimensional and non-normally distributed data has made it an attractive and powerful option for integrating different imagery sources and ancillary data sources into image classification workflows (Kloiber et al. 2015). For a more detailed description on RF theory and its parameter optimization, the reader is directed to studies done by Breiman (2001), Tian et al. (2016), Lin et al., (2010), Mutanga et al. (2012).

6.2.4 Support Vector Machine (SVM)

SVM is a machine learning distribution free classifier and does not encounter any over-fitting problem (Burges 1998; Cortes and Vapnik 1995). This supervised method is trained to find an optimal classification hyperplane by minimizing the upper bound of the classification error. There are two supporting hyperplanes on the boundaries of the data distribution and the data points on the edge of these hyperplanes are the support vectors of the algorithm. But the problem is all classes of an image are not linearly separable,

hence, SVM is optimized to search for a non-linear hyperplane in a multidimensional feature. This transformation is implicitly performed by applying kernel functions to the original data (Keramitsoglou et al. 2006). There are two commonly used functions on remotely sensed data, non-linear polynomial and radial basis function (RBF) kernels (Huang et al. 2002; Oommen et al. 2008). Seeing as all classes of an image are not linearly separable, SVM is optimized to search for a non-linear hyperplane in a multidimensional feature. This transformation is implicitly performed by applying kernel functions to the original data (Hornik et al. 2006). Two functions are commonly used on remote sensing data, non-linear polynomial and radial basis function (RBF) kernels (Huang et al. 2002; Oommen et al. 2008). In this study, eight bands of Worldview 2 were used for defining the space feature of SVM through a radial basis kernel function. We used RBF because a number of authors have found that the RBF outperforms the polynomial function(Kavzoglu and Colkesen 2009; Oommen et al. 2008; Waske and Braun 2009). In addition, this RBF is computationally fast and easy to implement as it requires only two tuning parameters, cost (c) and gamma (γ). More detailed information about SVM can be found in the literature by Hornik et al. (2006); Cortes and Vapnik (1995), and Mathur and Foody (2008a).

6.2.5 Feature Selection

Random Forest (Breiman 2001) is based on a similar technique to Classification and Regression Tree (CART), however, instead of a single tree (as in CART), RF grows a large number to trees to build a forest. Here accuracy can be measured based on Out-of-Bag(OOB) error, i.e. the mean square error (for prediction) and misclassification error (for classification) averaged over observations from all trees for which they have been OOB (Lu and Petkova 2014). Two variable selection approaches, Linear Vector Quantization (LVQ) and Recursive Feature Elimination (RFE) were used that are available in Caret packages in R. The Learning Vector Quantization algorithm (or LVQ) is an artificial neural network algorithm. Feature selection through LVQ that can reduce the dimensionality of the input variables, which can improve the accuracy of the method. LVQ suffers from the same curse of dimensionality in making predictions as K-Nearest Neighbours. To overcome this problem, we used another feature selection, called recursive feature elimination method (RFE).

6.2.6 Parameter optimization

RF works based on two tuning parameters, the number of trees in the ensemble (*centre*), and the number of variables randomly sampled at each node to be considered for splitting (*mtry*) (Peters et al. 2002). In principle, one should simultaneously optimize the number of trees (*ntree*) and mtry. Usually, the default number of trees (*ntree*) is 500, while the default value for the number of variables (*mtry*) is the square root of the total number of spectral bands used in the study (Breiman 2001). In our study, a grid-search approach

based on the accuracy was used to find the optimal combination of these two parameters (Tian et al. 2016). The grid search value for mtry varied from 1 to 8, while the grid search value for the ntree parameter varied from 500 to 2500 with an interval of 500. Then the optimal setting for RF parameters (*mtry* and *ntree*) was input into the R platform to retrieve the confusion matrices of accuracy. However, this computation process is intractable. We used the Random Forest, *e1071* library, and Caret (Kuhn 2008) packages in the R statistical software (R development core team 2016) to tune the parameter, variable selection and execute the machine learning regression method. For SVM, we used the Radial Basis Function (RBF) kernel that is composed of two parameters, cost (C) and gamma (γ). The C value is used for adjusting the error of misclassifying instants of the training dataset and γ is used to determine the width of the kernel.

6.2.7 Evaluation of performance

The Z-statistic (Congalton and Mead 1986) was used to assess the significance of the difference between resulting classifications obtained by different experiment sets based on original and reduced band numbers. The difference between the two classification results is considered to be significant at the 95% confidence level if the absolute value of the Z-statistic exceeded 1.96. The Z-statistics is calculated as below:

$$Z_{ab} = \frac{|K_a - k_b|}{\sqrt{var_{(Ka)} - Var_{(Kb)}}}$$

where Z_{ab} is the *Z*-statistic for comparison of classification *a* and *b*; K_a and K_b are the kappa coefficients of classifications *a* and *b*; and var(κ_a) and var(κ_b) are the asymptotic variances of *κa* and *κb*, respectively.

6.3 Results

6.3.1 Parameter optimization in RF

The results of Random Forest parameters (*mtry* and *ntree*) are shown in figure 6-1. The optimization was performed using the three calibration datasets and the accuracy. RF parameters (*ntree* and *mtry*) clearly affected the error of prediction. We varied both parameters across a wide range of values (number of trees: 500 to 2,500) and (number of random split variables at each node: 1 to 8). The overall accuracy for all three experiments was derived from 10-fold cross-validation of the calibration dataset. Using two randomly selected input variables per node delivered better accuracy than using 4, 6 or 8 variables for experiment 2 and 3. A high number of trees (>2000) produced the lowest

prediction errors. These selected ntree and mtry values were further used to develop the confusion matrix from the test dataset.

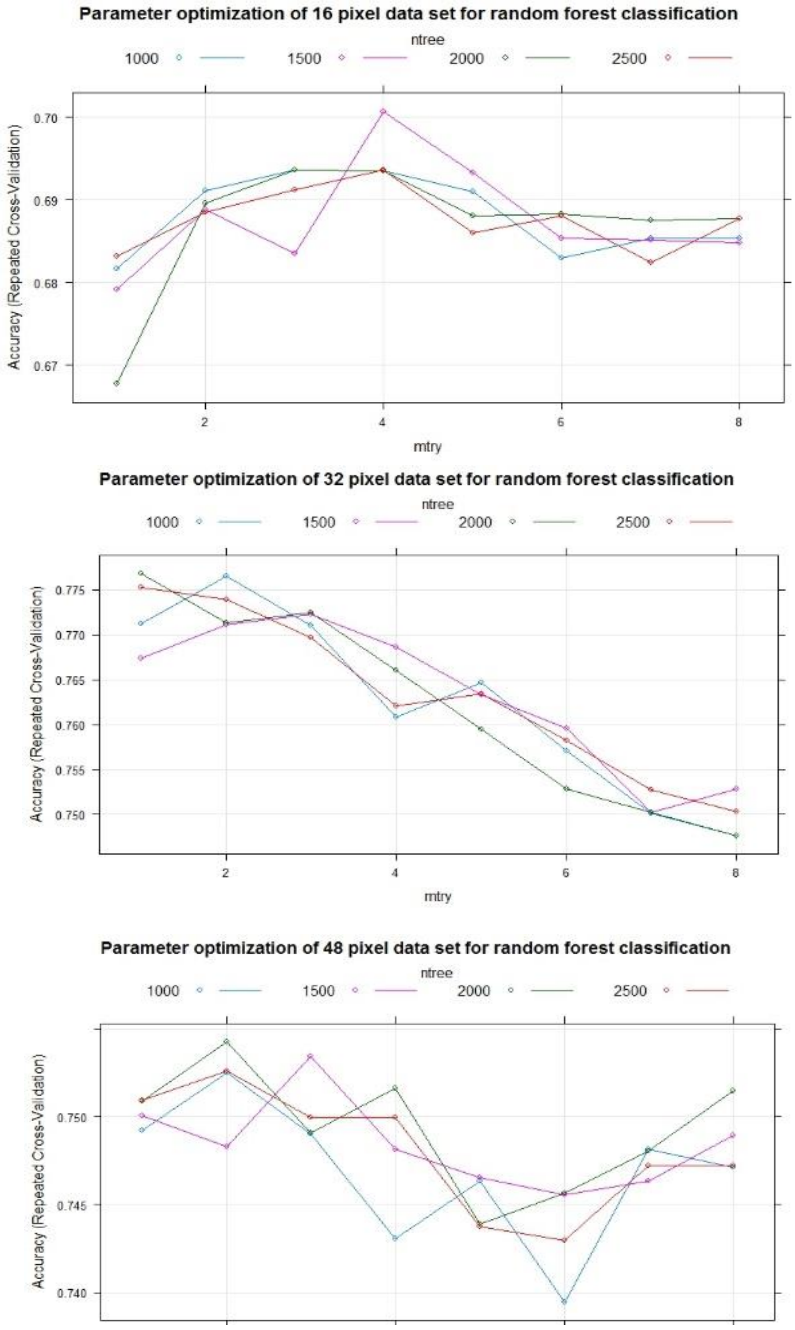

Figure 6-1: . Effect of the number of trees (ntree) and the number of random split variables at each node (mtry) on the overall accuracy of the RF classification using the 10 folds cross validation on 8 bands of WorldView-2.

Results from the parameter optimization indicated different ntree and mtry values for each of the three experiment sets. The default mtry value of 2 combined with a ntree value of 500 produced the lowest OOB error rate (23.44%) for the 32 pixels experiment set (table 6-3). When this dataset was tuned with optimized parameter (ntree = 1000, mtry = 2), the OOB error rate reduced to 22.66 % that is the lowest OOB error rate within the three experiment sets. The study of error estimates for bagged classifiers in (Breiman, 2001), gives empirical evidence to show that the out-of-bag estimate is as accurate as using a test set of the same size as the training set. However, when three calibration sets were validated with an independent test data set, RF classification accuracy reduced when sample size was proportionally reduced.

Table 6-3: OOB error rate before and after parameter optimization for Random Forest classification.

Calibration set	OOB error %	Parameter tuning	After tuning OOB%
48 pixels	24.44	ntree= 2000, mtry =2	23.96
32 pixels	23.44	ntree= 1000, mtry =2	22.66
16 pixels	31.25	ntree= 1500, mtry =4,	30.47

6.3.2 Relationship between pixels number of the dataset and classification accuracy

Table 6-4: Random Forest confusion matrix for 48 pixels dataset

	CA	DI	MA	GR	PH	SP	WA	WS	Total	EO	UA
CA	29	1	9	1	0	0	4	1	45	35.5	64.4
DI	0	41	0	1	9	0	2	4	57	28.1	71.9
MA	11	0	33	2	0	0	1	0	47	29.7	70.2
GR	3	0	5	42	1	2	3	0	56	25	75
PH	0	2	0	1	37	2	1	0	43	13.9	86.0
SP	3	1	0	1	1	44	0	0	50	12	88
WA	2	1	1	0	0	0	34	1	39	12.8	87.1
WS	0	2	0	0	0	0	3	42	47	10.6	89.3
Total	48	48	48	48	48	48	48	48	384		
EC	39.58	14.58	31.25	12.5	22.9	8.33	29.16	12.5	O A = 0.79		
PA	60.41	85.41	68.75	87.5	77.1	91.6	70.83	87.5	Kappa = 0.76		

Table 6-5: Random Forest Confusion matrix for the 32-pixel dataset.

	CA	DI	MA	GR	PH	SP	WA	WS	Total	EC	UA
CA	14	0	6	1	0	0	2	0	23	39.1	60.9
DI	0	28	0	1	5	0	5	2	41	31.7	68.3
MA	11	0	19	2	0	0	2	0	34	44.1	55.8
GR	2	0	6	27	1	3	0	0	39	30.7	69.2
PH	0	2	0	1	25	2	1	0	31	19.4	80.6
SP	3	0	0	0	1	26	0	1	31	16.1	83.8
WA	2	1	1	0	0	1	22	1	28	21.4	78.6
WS	0	1	0	0	0	0	0	28	29	3.4	96.5
Total	32	32	32	32	32	32	32	32	256		
EC	56.25	12.5	40.62	15.62	21.87	18.75	31.25	12.5	OA = 0.74		
PA	43.75	87.5	59.38	84.37	78.12	81.25	68.75	87.5	Kappa = 0.70		

Table 6-6: Random Forest confusion matrix for 16-pixel dataset.

	CA	DI	MA	GR	PH	SP	WA	WS	Total	EC	UA
CA	4	1	1	0	0	2	0	2	10	60	40
DI	0	11	0	0	5	0	1	0	17	35.29	64.7
MA	9	0	15	0	0	0	1	0	25	40	60
GR	1	0	0	14	0	0	2	0	17	17.65	82.3
PH	0	2	0	0	11	0	0	0	13	15.38	84.6
SP	0	1	0	2	0	14	1	0	18	22.22	77.7
WA	0	0	0	0	0	0	11	1	12	8.33	91.6
WS	2	1	0	0	0	0	0	13	16	18.75	81.2
Total	16	16	16	16	16	16	16	16	128		
EO	75	31.25	6.25	12.5	31.25	12.5	25	18.75	O.A = 0.72		
PA	25	68.75	93.75	87.5	68.75	87.5	75	81.25	Kappa = 0.68		

Key for (table 4,5 & 6): CA=*Casuarina*, DI= Dieback *Casuarina*, MA= Mangrove, GR= Grass(perrenial), PH = *Phragmites australis*, SP=*Sporobolus virginicus*, WS = Wetland Soil. PA = Producer Accuracy, UA = User Accuracy, EC = Error of commission, EO= Error of Omission.

As sample size increased, error in general decreased (table 6-3). Therefore, increasing sample size, lead to improved classification accuracy. This is for both the calibration set with cross-validation (table 6-3) and the independent test dataset (table 6-4,6-5 and 6-6). From confusion matrix table of independent test dataset, it is clear that classification accuracy dropped from 79% to 72% when pixels number is reduced from 48 to 16 per class (table 6-4,6-5 and 6-6).

6.3.3 Parameter optimization in SVM

The tune function for parameter optimization resulted in the same gamma value for all experiment sets (table 6-7) except the 48-pixel dataset (Figure 6-2). The cost parameter varied from 1.00 to 2.00 for all three of the calibration datasets. Tuned parameters also improved the accuracy of all three calibration datasets (Table 6-7). A significant observation was found for the16 pixels dataset where accuracy improved from 0.78 to 0.86 when tuned parameters were applied (table 6-7).

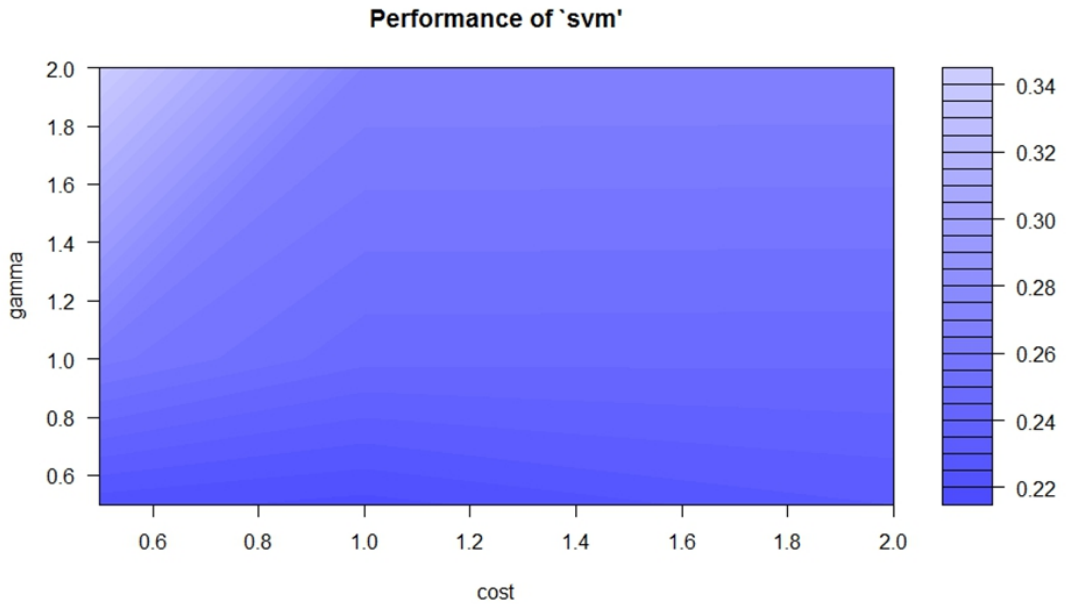

Figure 6-2: Optimization of SVM parameters (cost and gamma) based on the tune function. Lighter blue indicates greater accuracy.

Table 6-7: Tuning of SVM parameter

Experiment set	Accuracy	Tuned parameter	Accuracy	P value
16 pixels set	0.78 Kappa= 0.75	Cost = 1, Gamma = 0.5	0.86 Kappa = 0.84	< 0.0000001
32 pixels set	0.81 Kappa = 0.78	Cost = 2, Gamma= 0.5	0.85 Kappa = 0.83	< 0.0000001
48 pixels set	0.79 Kappa = 0.75	Cost = 2, Gamma = 2	0.82 Kappa= 0.80	< 0.0000001

Like RF, the SVM classifier was also trained using three training sets, respectively, and then those calibrated classifiers were applied to classify the testing set. Figure 6-2 and Table 6-7 (SVM calibration) illustrate the changes of accuracy against training set size by this method. The accuracy of wetland classification results by SVM is significantly higher than that of RF for all training sets. Whilst the cross-validation of RF provided 70% accuracy for the 16-pixel data set (Table 6-3), SVM increases to 86% (Table 6-7).

6.3.4. The relation between pixels number of the training dataset and classification accuracy for SVM

Table 6-8: SVM Confusion matrix for 16 pixels dataset

	CA	DI	MA	GR	PH	SP	WA	WS	Total	EO	UA
CA	3	0	1	0	0	1	0	0	5	40	60
DI	0	12	0	0	4	0	0	0	17	23.52	76.47
MA	9	0	15	0	0	0	1	0	23	34.78	65.21
GR	0	0	0	14	0	0	2	0	16	12.5	87.5
PH	0	3	0	1	12	0	1	1	17	29.41	70.58
SP	0	0	0	1	0	15	0	0	16	6.25	93.75
WA	0	0	0	0	0	0	10	0	12	16.6	83.33
WS	4	1	0	0	0	0	2	15	22	31.8	68.18
Total	16	16	16	16	16	16	16	16	128		
EC	81.25	18.75	6.25	12.5	25	6.25	37.5	6.25	OA = 0.75		
PA	18.75	81.25	93.75	87.5	75	93.75	62.5	93.75	Kappa = 0.71		

Table 6-9: SVM Confusion matrix for 32 pixels dataset

	CA	DI	MA	GR	PH	SP	WA	WS	Total	EC	UA
CA	15	0	8	3	0	0	2	0	28	46.42	53.57
DI	0	26	0	1	3	0	3	2	35	25.71	74.28
MA	9	0	17	1	0	0	1	0	28	39.28	60.71
GR	3	0	6	26	1	2	0	0	38	31.57	68.42
PH	0	5	0	1	27	2	1	0	36	25	75
SP	3	0	0	0	1	28	0	0	32	12.5	87.5
WA	2	1	1	0	0	0	23	1	28	17.85	82.14
WS	0	0	0	0	0	0	2	29	31	6.45	93.54
Total	32	32	32	32	32	32	32	32	32		
EC	53.12	18.75	46.87	18.75	15.62	12.5	28.12	9.37	256	OA = 0.75	
PA	46.87	81.25	53.12	81.25	84.37	87.5	71.87	90.62		Kappa = 0.71	

Table 6-10: SVM Confusion matrix for 48 pixels dataset

	CA	DI	MA	GR	PH	SP	WA	WS	Total	EC	UA
CA	27	0	8	3	0	0	2	1	41	34.14	65.85
DI	1	40	0	1	7	0	4	2	55	27.27	72.72
MA	11	0	35	1	0	0	1	0	48	27.08	72.91
GR	0	0	4	41	2	2	2	0	51	19.60	80.39
PH	0	4	0	1	38	3	1	0	47	19.14	80.85
SP	3	1	0	1	1	43	1	1	51	15.68	84.31
WA	4	0	1	0	0	0	32	0	37	13.51	86.48
WS	2	3	0	0	0	0	5	44	54	18.51	81.48
Total	48	48	48	48	48	48	48	48	384		
EO	43.75	16.66	27.08	14.58	20.83	10.41	33.33	8.33		OA = 0.76	
PA	56.25	83.33	72.91	85.41	79.16	89.58	66.66	91.66		Kappa = 0.72	

Based on the optimized parameters applied to the test data set, it was demonstrated that Overall accuracy (OA) did not change when pixel numbers were reduced from 48 pixels to 16 pixels per class (Tables 6-8,6-9 & 6-10).

6.3.5 Feature selection and its effect on accuracy

We examined the explanatory power of the 8 spectral bands of Worldview 2 data in the classification of the 6 plant species with other two land cover classes using the following measures of variable importance: Learning Vector Quantization and Recursive feature elimination. At first, the measures of variable importance were evaluated in combination with the pairwise correlations shown in figure 4.

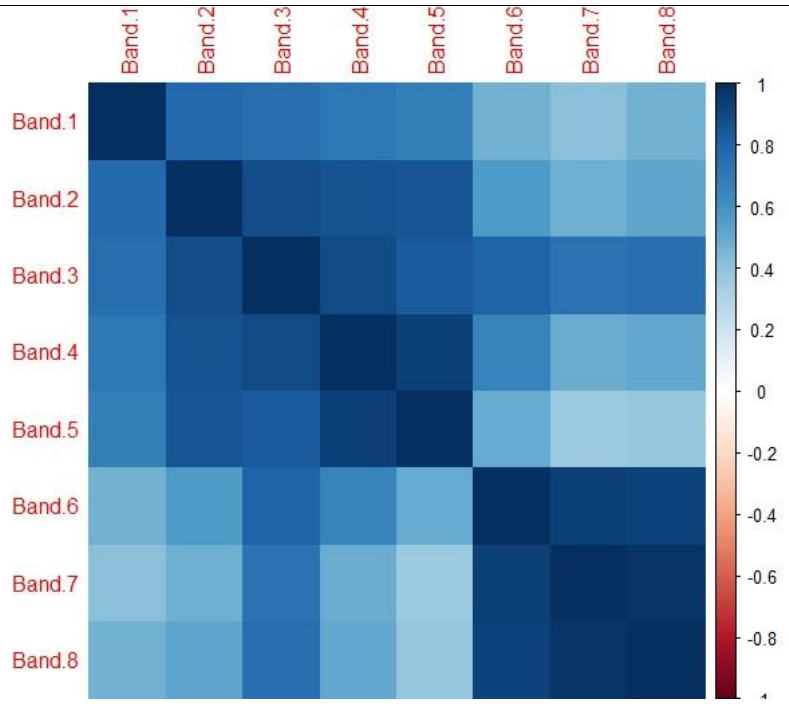

Figure 6-3: Pairwise correlation between different bands of Worldview 2 data based on 32 pixels data set.

Recursive Feature Elimination (RFE) (figure 6-5) selected 6 variables that were highly important and were similar to the selection of the LVQ model (figure 6-5) and covariance matrix (figure 6-4) information. Based on our findings coastal band and yellow are less important. Although band 7 (Near Infra-Red 1) and band 8 (Near Infrared 2) are highly correlated (figure 6-4), still they are important for vegetation and wetland soil classification (Tarantino et al. 2012). Coastal blue and yellow bands are mainly used for open sea water and sandy coastal area classification (Tarantino et al. 2012) that are absent in our study area. For all vegetation classes, red-edge (Band 5) is very important due to its sensitivity to the vegetation chlorophyll content and canopy structure (Mutanga et al. 2012).

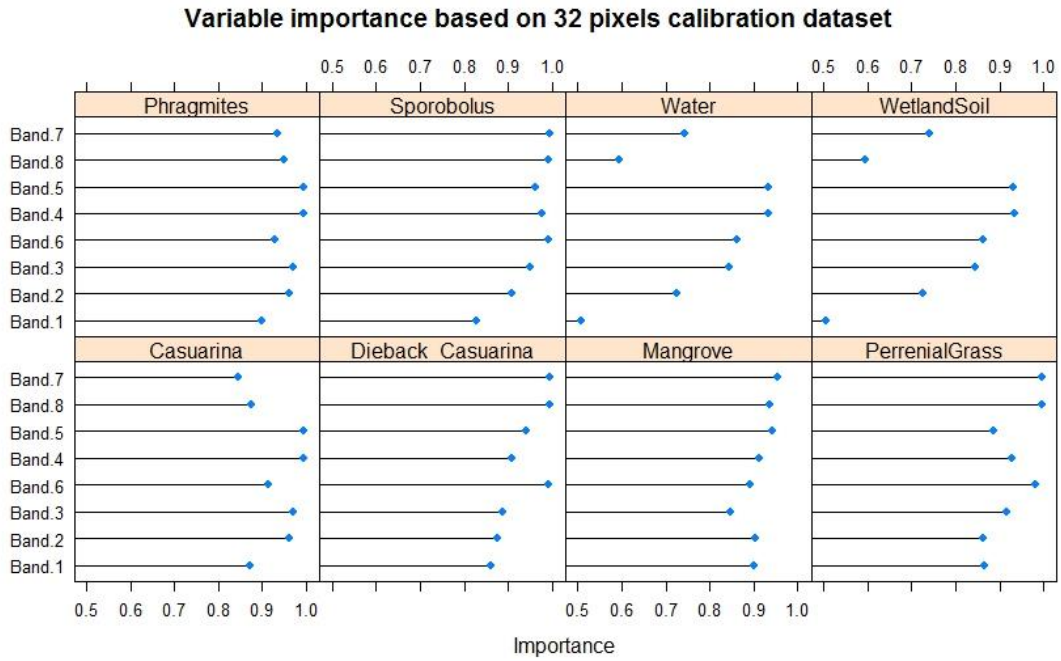

Figure 6-4: Linear Vector Quantization (LVQ) feature selection based on 32 pixels set calibration data.

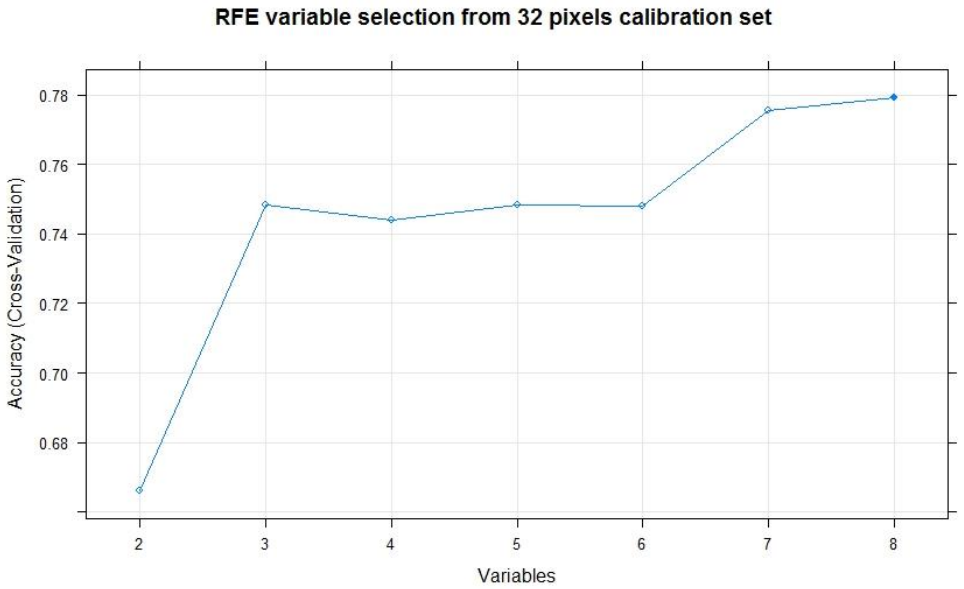

Figure 6-5: Figure 6. Recursive Feature Elimination (RFE) for feature selection from Worldview 2 data.

By varying the sample size and running classifications (with repeated cross-validation), it was evident that when dimensionality was reduced and only the most important variables were considered, OOB error % percentages were much more similar to the OOB error percentages of the "all variables" classification (Table 6-3 and 6-11). However, there was no change in the kappa when SVM was applied for selected features only (table 6-12).

Table 6-11: OOB error rate before and after parameter optimization for RF classification (for the selected feature only).

Calibration set	OOB error %	Parameter tuning	After tuning OOB%
48 pixels set	23.70	ntree = 1000, mtry =2	22.66
32 pixels set	25.78	ntree= 2000, mtry=2	24.22
16 pixels set	30.47	ntree= 2000, mtry = 4	28.91

Table 6-12: SVM accuracy after feature selection

Experiment set	Parameter	Accuracy	Kappa
16 Pixel set	C = 1, γ = 0.5	0.76	0.72
32 Pixel set	C = 2, γ = 0.5	0.76	0.72
48 Pixel set	C = 2, γ = 2	0.75	0.72

Classification matrices originating from these two different datasets (optimal six bands and original eight bands) are presented in the Figure of 6-6,6-7, 6-8 and 6-9. Figures 6-7 and 6-9 show the effect of feature selection on three experiment sets for R classification. In both cases, the classification trend is similar and hence there is no significant difference between the user's accuracy in 8 different classes. When the number of bands number is reduced from 8 to 6 based on feature selection method, user's accuracy and overall accuracy did not change significantly (Figure 6-7 and 6-9). This demonstrates the importance of feature selection for RF and how it preserves the accuracy level even with the reduced number of bands.

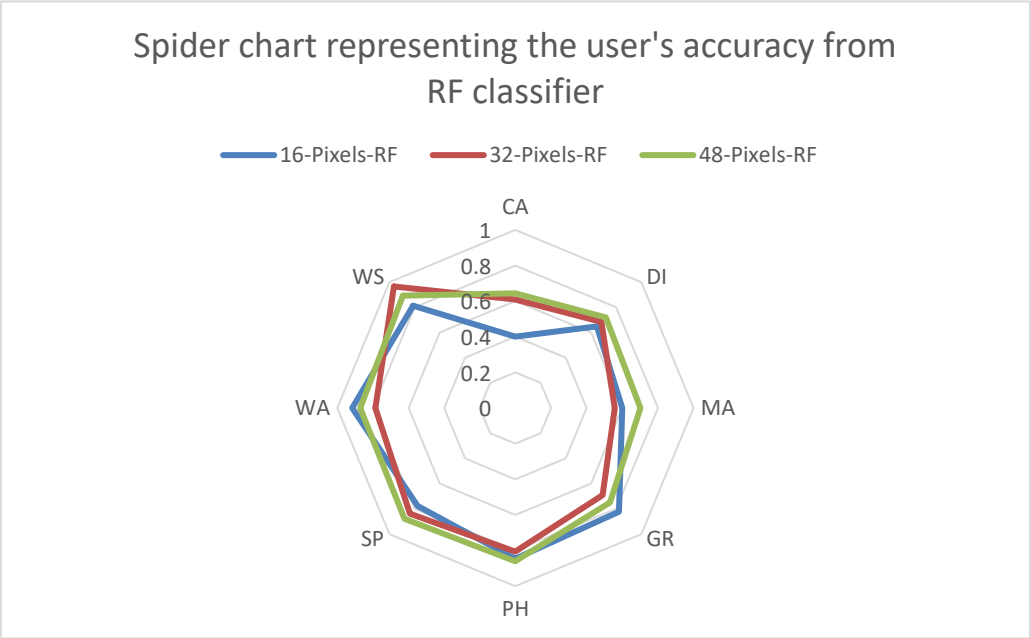

Figure 6-6: Spider charts representing the user accuracies for the Random Forest classification based on two different band combinations for all bands of Worldview-2

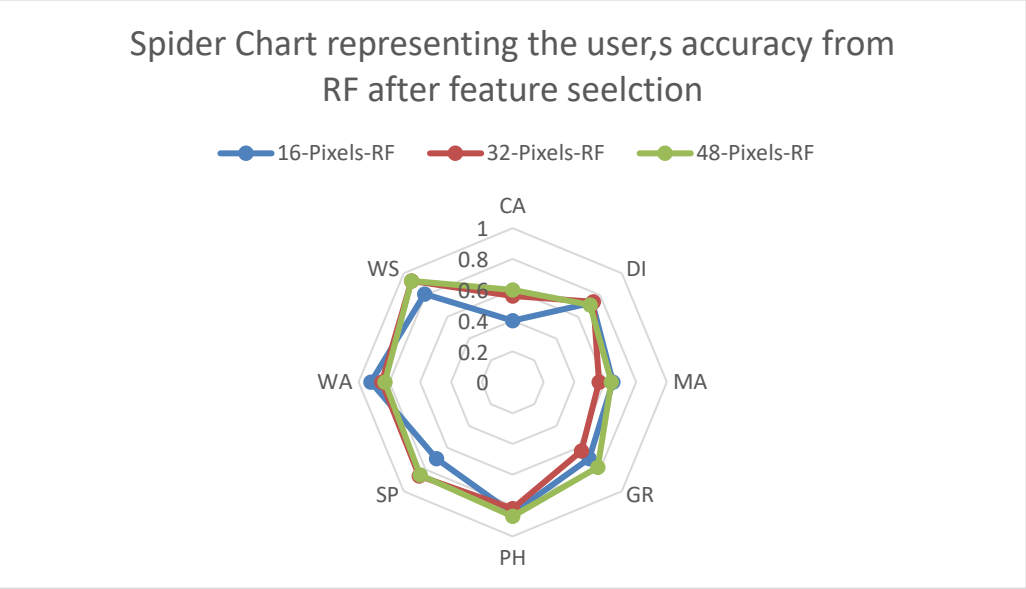

Figure 6-7: Spider charts representing the user accuracies for the Random Forest classification based on two different band combinations for only selected bands of Worldview-2 chosen by recursive feature elimination (RFE).

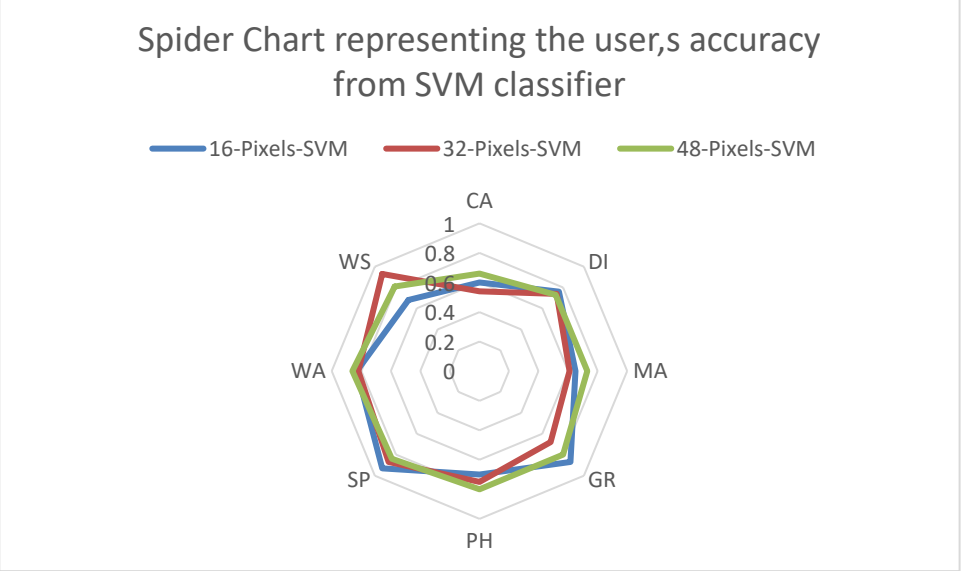

Figure 6-8: Spider charts representing the user accuracies for the Support Vector Machine (SVM) classification based on two different band combinations, (top) for all bands of Worldview 2.

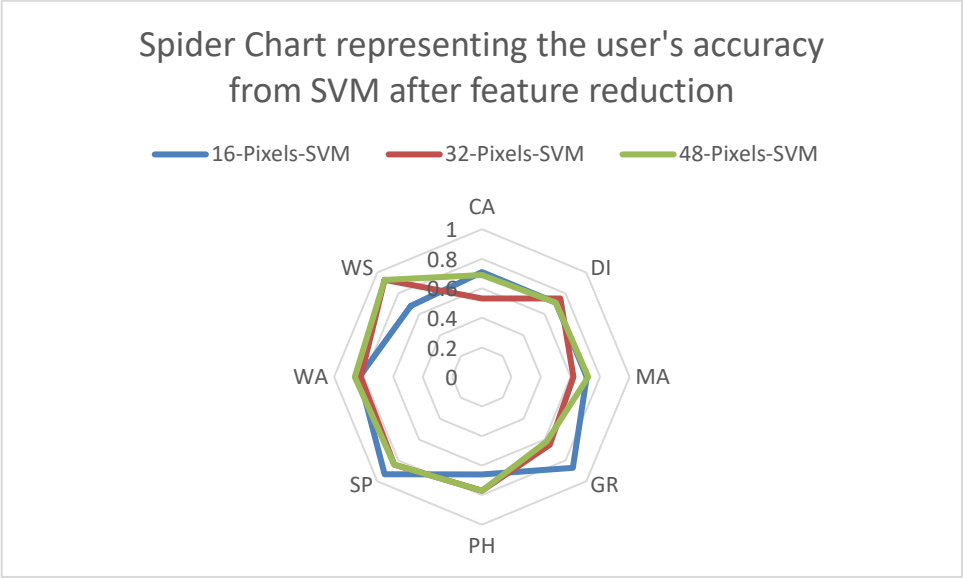

Figure 6-9: Figure 8: Spider charts representing the user accuracies for the Support Vector Machine (SVM) classification based on two different band combinations, (top) for all bands of Worldview 2 and (bottom) only selected bands of Worldview 2 chosen by recursive feature elimination (RFE).

Comparing the two charts (Figure 6-8 & 6-9), the effect of feature selection is very clear. User accuracies have not changed significantly when only selected six bands were used for classification.

Table 6-13: Maximum Likelihood Classifier (MLC) for the 48-pixel dataset (confusion matrix) with a summary of 32 and 16-pixel datasets.

	CA	DI	MA	GR	PH	SP	WA	WS	Total	EC	UA
CA	27	0	20	3	0	0	3	1	54	50	50
DI	0	42	0	0	4	0	7	2	55	23.64	76.3
MA	16	0	19	2	0	0	0	0	37	48.65	51.3
GR	1	0	7	35	2	5	2	0	52	32.69	67.3
PH	1	6	0	4	41	5	2	1	60	31.67	68.3
SP	1	0	0	1	1	38	1	1	43	11.63	88.3
WA	1	0	0	3	0	0	31	0	35	11.43	88.5
WS	1	0	2	0	0	0	2	43	48	10.42	89.5
Total	48	48	48	48	48	48	48	48	384		
EO	43.75	12.50	60.42	27.08	14.58	20.83	35.42	10.42		OA = 0.71	
PA	56.25	87.50	39.58	72.92	85.42	79.17	64.58	89.58		Kappa= 0.67	
32 pixels data set								OA =0.69		Kappa=0.65	
16 pixels data set								OA =56.25		Kappa= 0.50	

We also evaluated the traditional Maximum Likelihood Classifier (MLC) and it was found that maximum accuracy obtained from MLC was 71% from the 48-pixel dataset. However, it is evident that with the increase of the sample size classification accuracy greatly improved. That is the traditional nature of MLC and need a large sample size for improved accuracy.

6.3.6 Evaluation of performances

An overview of all classification results based on significance level (Z –statistic) is provided in Table 6-14. The following models are compared: Random Forest versus Support Vector Machine (All possible combinations within three experiment sets, and all bands versus reduced bands (selected features).

Table 6-14: Z-statistics among classification result

Pair comparison	Z -statistics
Worldview 2 (8 spectral bands)	
16 Pixels – RF vs 32 Pixels -RF	0.414
32-Pixels RF vs 48-Pixels RF	1.562
48-Pixels RF vs 16 Pixels RF	1.980*
16 Pixels-SVM vs 32 Pixels-SVM	NS
32 Pixels SVM vs 48 Pixels SVM	NS
48 Pixels SVM vs 16 Pixels SVM	NS
16 Pixels RF vs 16 Pixels SVM	0.521
32 Pixels RF vs 32 Pixels SVM	0.248
48 Pixels RF vs 48 Pixels SVM	1.264
Optimal bands (after feature selection)	
16 Pixels – RF vs 32 Pixels -RF	0.387
32-Pixels RF vs 48-Pixels RF	1.562
48-Pixels RF vs 16 Pixels RF	1.980*
16 Pixels-SVM vs 32 Pixels-SVM	NS
32 Pixels SVM vs 48 Pixels SVM	NS
48 Pixels SVM vs 16 Pixels SVM	NS
16 Pixels RF vs 16 Pixels SVM	0.544
32 Pixels RF vs 32 Pixels SVM	0.212
48 Pixels RF vs 48 Pixels SVM	0.677

6.4 Discussion

6.4.1 Performance of non-parametric classifiers

The non-parametric RF and SVM classifiers performed almost equally as there is no significant difference between the performances of these two. Comparing RF and SVM regarding their requirements on the data, we see the following advantages of RF for wetland classification: in contrast to MLC, the non-parametric SVM and RF classifiers do not make any assumptions about data distribution. Therefore, unlike MLC, they can handle multi-modal data distributions. In our study, the classes were mostly normally distributed leading to similar results with RF and SVM. However, this is not always the case for image analysis with remote sensing data. There are a couple of advantages of RF for wetland species classification. Firstly, RF does not require that the classes have a common covariance matrix, which often is not the case in wetland species classification and hence limits the use of parametric classifiers. In addition, RF provides a reliable measure of variable importance, i.e., RFE (Recursive Feature Elimination) that is very helpful for feature selection, as demonstrated in the paper.

This study differs from Carle et al. (2014) who suggested that MLC performed better than SVM when data are normally distributed. Although our data were normally distributed, still SVM and RF both performed better than MLC. The main reasons are our high spatial resolution data allowing a small training set size. Jin et al. (2005) demonstrated that MLC has difficulty dealing with texture features of high-resolution images, especially in small training set size situations. In addition, MLC suffers from the Hughes effect or the curse of dimension (Hughes 1968; Pal and Mather 2003) . This means that with a fixed number of training samples, performance declines with increasing dimensions due to the decreased reliability of estimates of statistical parameters needed to calculate the probability (Oommen et al. 2008).

We used a nonlinear RBF kernel function to perform SVM classification. A nonlinear kernel is a robust method to solve inseparability problems that may be found in the heterogeneous classes of wetland ecosystem (Adam et al. 2014). The findings of our study obtained from two classifiers are consistent with Pal (2005), Sesnie et al. (2010) and Adam et al. (2014). However, Pouteau et al. (2012) and Nitze et al. (2012) found incomparable classification accuracy from the two classifiers. Both RF and SVM were unable to improve the user accuracy for Grey-mangrove and Swamp-oak species due to the spectral similarities between them. Similar problems also observed by Duro et al. (2012) when they used high spatial resolution data for pixel-based classification for a heterogeneous landscape. Although different authors (Duro et al. 2012; Schuster et al. 2012) recommended a post-classification process to improve the accuracy, we did not use the post-classification process as our study aimed at pixel-based classification(Adam et al. 2014).

6.4.2 Training sample size

The literature often suggests that the size of the training set required is a function of spectral wavebands used and generally a sample comprising at least 30 times the number of spectral bands used is required in the analysis process (Mathur & Foody, 2008). Moreover, it has been demonstrated that training set size has a positive relationship with the classification accuracy(Arora and Foody 1997; Pal and Mather 2003). That is why conventional classification method (like MLC) require a large training set distributed throughout the study area. With this approach, Worldview-2 data requires at least 240 pixels for each class to run a conventional MLC algorithm. In contrast, only 15% (16 pixels for each class) of the possible required training sample (240 pixels) provided up to 75% accuracy from machine learning classifications. This accuracy is acceptable not only based on cost-benefit analysis but the feasibility test as well. For wetlands in many settings, it has been difficult and expensive to collect at least 240 pixels of training data for each class to test MLC performance. The findings of the current study are supported by Mathur and Foody (2008) when they reduced training size from 450 pixels to 130 pixels and classification accuracy reduced only 1.34% (statistically insignificant). In their study, SVM provided 1.34% less accuracy (90.66%) than MLC (92.00%), Mathur and Foody (2008a) concluded in favour of the superiority of SVM due to small training size requirements and the relatively low cost of analysis.

From the findings shown in Figure 6-2 and Table 6-7 (cross-validation for training data), it is evident that SVM is quite tolerant to training set size, having a super generalization capability even with a small set of training samples. This is also supported by Song et al. (2012) who received similar results using Artificial Neural Network. The main reason is that only support vectors that consist of a few margin samples are considered by SVM for classification; other samples make no more contribution to classification (Foody and Mathur 2004). That is why a small set of samples can be separated into a few groups to train the SVM for obtaining its optimal parameters by the cross-validation method.

The results of this study also demonstrate that RF image classification is sensitive to training data sample size. A larger training sample size produced a lower OOB, therefore increased accuracy. Unlike SVM, larger training sample sizes are recommended to improve classification accuracy and stability with the RF classifier. Our findings are consistent with the findings of Millard and Richardson (2015) as they also revealed that RF is sensitive to training sample size and larger training size increases the overall accuracy. One of the limitations of RF is that running iterative classifications using the same training and input data was found to produce different classification results. Therefore parameter optimization based on cross-validation was performed before testing the model performance. These optimized parameters also controlled the overfitting of the final model.

Although Millard and Richardson (2015) found that RF variable importance measures and rankings varied with iterative classifications, our variable importance measures provided consistent results from both of the feature selection methods.

6.4.3 Equal sample allocation for training and validation datasets

The process of training data creation for remote sensing image classification is a crucial issue and involve methodological choices that provide trade-offs in data quality (class representativeness) and quantity (size of the data). Therefore, training data points where field validation has been done result in high certainty in the training dataset with a minimum level of bias. But as mentioned before, these points are often difficult to obtain and therefore there may be a tendency for researchers to use fewer training data sample points. This may be especially the case for wetland environments, time and access constraints may require researchers to obtain classifications with minimal training sample points. Many researchers conduct image classification from training sample points collected from imagery (i.e. area of interest/polygon) that is quick and easy. This method produces highly clustered training sample points (within a polygon) with inherently high spatial autocorrelation (Millard and Richardson 2015). Although this method produces greater accuracy (Rasel et al. 2016a), a high level of bias is also an issue when training samples are highly autocorrelated (Millard and Richardson 2015). In this study, individual pixels collected from the field were randomly distributed and independent from training sets that help to avoid bias (Song et al. 2012). Equal numbers of samples were allocated and reduced gradually to observe the effect on accuracy. Although the performance of SVM based on small sample sizes has already demonstrated, equal number of sample distribution provided the robustness of the model. This was ensured as each time calibration models were validated with an independent test dataset from eight classes. RF draws its validation data from the training dataset and provides OOB error for assessment. In this study, equal amounts of samples (gradually reduced) controlled the overfitting of the RF model as each time the RF subset originated from the independent data that equally represented each class and were not spatially autocorrelated.

6.4.4 Feature selection and accuracy

Our findings show that variable reduction should be performed to obtain the optimum classification. Although our dataset was limited to 8 spectral bands, six optimum bands did not affect the overall accuracy. This demonstrates the power of the feature selection methods. Not only this, Millard and Richardson (2015) found that when high dimensional datasets were used, classification results were noisy. Other previous research (Millard and Richardson 2015) has shown that although RF can handle high dimensional data, classification accuracy remains relatively unchanged when only the most important predictor variables are used. Similarly, we demonstrated that removing less important variables from the original variables did not change the accuracy. Removing variables of

lesser importance also improved the stability in classification, based on the model parsimony test. This was because fewer variables maintained the performance of models and require less computational work than high dimensional data.

6.5. Conclusion and recommendation

Two types of advanced non-parametric classifiers, Random Forest (RF) and Support Vector Machine (SVM), were tested for saltmarsh dominant wetland classifications using 8 band Worldview 2 images. This work utilised equal sample allocation strategies and reduced the sample sizes proportionally (100% to 33%) to assess the effect of sample size on calibration and validation of the classifier. All classifiers produced acceptable classification results with all sizes of training set data but the RF performed slightly better than SVM models according to kappa analysis with testing data. Only RF showed significant changes when the sample size was reduced from the 100% level to 33% and overall accuracy dropped from 79% to 72 %. But there were no significant changes in the accuracy when samples size equally dropped from 100% to 33% for SVM. Based on equal sample allocation strategies, only 15% of the required samples provided for 75% of the overall accuracy from the SVM. However, there was no significant difference among the all possible combinations of three experiments (100 %, 66% and 33% of the original samples) originated from two classifiers based on equal sample distribution scheme. Two feature selection methods provided by Classification and Regression Training (CARET) and Random Forest (RF) packages in R environment were evaluated to test the importance of spectral bands. Learning Vector Quantization (LVQ) and Recursive Feature Elimination (RFE), both selected six important features. No significant differences were observed when these six important features were selected instead of all features for classification accuracy. Although two near-infrared bands are highly correlated, they are very important for classification with a red edge and red bands. This accuracy is acceptable and is comparable to other previous studies where multispectral and airborne hyperspectral data have been used. However, in terms of data acquisition cost, reference data collection and processing cost, this study showed a promising approach for broad management applications for saltmarsh community that are ecologically endangered, and demand a special attention.

Chapter 7

Saltmarshes biomass modelling using Random Forest (RF) and Support Vector Machine (SVM) regressions from multispectral data

This chapter is based on

The article that has been submitted to the International Journal of Digital Earth. Based on the first submission, reviewers recommended for publication with a major revision. Therefore, this chapter is the revised format that has been submitted (Ist revision) to the International Journal of Digital Earth (IJDE).

Abstract

Assessing large-scale plant productivity of coastal marshes is essential to understand the resilience of these systems to climate change. Two machine learning approaches, Random Forest (RF) and Support Vector Machine (SVM) regression were tested to estimate biomass of a common saltmarshes species, salt couch grass (Sporobolus virginicus). Reflectance and vegetation indices derived from 8 bands of Worldview-2 multispectral data were used for four experiments to develop the biomass model. These four experiments were, Experiment-1: 8 bands of the Worldview-2 image, Experiment-2: a Possible combination of all bands of Worldview-2 for Vegetation Indices, Experiment-3: Combination of bands and Vegetation Indices, Experiment-4: Selected variables derived from experiment-3 using variable selection methods. The main objectives of this study are (i) to recommend an affordable low-cost data source to predict biomass of a common saltmarshes species (ii) to suggest a variable selection method suitable for multispectral data (iii) to assess the performance of RF and SVM for the biomass prediction model. Cross-validation of parameter optimizations for SVM showed that the optimized parameter of ε-SVR failed to provide a reliable prediction. Hence, v-SVR was used for the SVM model. Among the different variable selection methods, Recursive Feature Elimination (RFE) selected a minimum number of variables (only 4) with an RMSE of 0.211 (kg/m2). Experiment-4 (only selected bands) provided the best results for both of the machine learning regression methods, RF (R^2= 0.72, RMSE= 0.166 kg/m2) and SVR (R^2= 0.66, RMSE = 0.200 kg/m2) to predict biomass. When a 10-fold cross validation of the RF model was compared with a 10-fold cross validation of SVR, a significant difference (p= <0.0001) was observed for RMSE. One to one comparisons of actual to predicted biomass showed that RF underestimates the high biomass values, whereas SVR overestimates the values; this suggests a need for further investigation and refinement.

Keywords: *Worldview-2, Salt couch, Spectral band, Vegetation Indices, Variable Selection*

7.1 Introduction

There is a crucial need to quantify large-scale plant productivity (i.e. above ground biomass) in coastal marshes for a better understanding of marsh resilience against sea level rise (Schile et al. 2014; Swanson et al. 2014). Aboveground biomass (AGB) is an important input parameter to develop models that forecast how coastal marsh elevations will respond to sea level rise (Morris et al. 2002; Swanson et al. 2014) . Moreover, AGB can be used to estimate belowground biomass (Gross et al. 1991) and soil organic carbon (Rasel et al. 2017). But this AGB estimation is labour intensive and not feasible at large spatial extents. To complement this, remotely sensed data are utilized to map vegetation types and provide better estimates of plant production (Goetz and Dubayah 2011). Combining field surveys and satellite observations for estimating AGB of wetland vegetation will help further the applications of belowground biomass and soil organic carbon (Rasel et al. 2017) and, indirectly, contribute to the climate change model of sea level rise.

Since the 1980s there have been some successful application of remote sensing for saltmarsh biomass (Gross et al. 1987; Hardisky et al. 1983; Hardisky et al. 1984) using Normalized Difference Vegetation Index (NDVI). Jensen et al. (2002) explained up to 70% of data variance when they used high-resolution image to map marsh biomass. Recently, the application of a red-edge band in an NDVI type index estimated biomass of wetland vegetation with relatively lower error (13% of observed mean biomass) (Mutanga et al. 2012). However, there are some major limitations of NDVI that have already been revealed by other studies (Gao et al. 2000; Jackson and Huete 1991; Kaufman et al. 1992; Middleton 1991; Qi et al. 1995; Sellers 1985; Tucker 1977). Several studies recommended using narrow bands of hyperspectral data to overcome the limitations (Blackburn 1998; Thenkabail et al. 2000). Other studies (Chen et al. 2009; Mutanga and Skidmore 2004a), using hyperspectral data, estimated biomass at full canopy cover with high accuracy. However, there are some crucial limitations of hyperspectral data that we experienced for our study area. First of all, the high cost to purchase aerial data if it is available. If data are unavailable for a specific study area, new acquisition is subject to season, sun illumination, weather conditions, flight schedules and aviation restrictions. In addition, time and processing costs of high dimensional (due to its narrow continuous bands) data are severe obstacles to general users, (e.g. wetland managers).

To overcome these limitations, Worldview-2 with its higher spatial (1.84 m) and spectral (8 bands) resolution is seen as a tradeoff between the advantages of multispectral resolution satellite data and hyperspectral data (Mutanga et al. 2012; Rasel et al. 2016b). WorldView-2 contains a special band named red-edge that facilitates vegetation related predictions. This is the region of abrupt change in the leaf reflectance between 680 and 780 nm due to the combined effects of strong chlorophyll absorption in red wavelengths and high reflectance in the NIR wavelengths(Horler et al. 1983) . Moreover, Worldview-2

data reduced the redundancy of information that is a problem of hyperspectral data (Omar 2010; Sridharan 2010). Recently some authors (Byrd et al. 2014; Mutanga et al. 2012) have used WorldView-2 imagery for estimating biomass in the wetland area. However, Byrd et al. (2014) predicted biomass with Landsat 7 (R^2 = 0.56, RMSE = 20.9%) that was slightly better than the results of Worldview 2 ((R^2 = 0.45, RMSE = 21.5%) for a mixed species model for two saltmarsh species. So there is a research gap for a species-specific model for saltmarsh biomass estimation from a multispectral platform. Moreover, Byrd et al. (2014) also recommended considering the tradeoffs between cost, spectral information and the high spatial resolution of most commercial satellite imagery, which can identify within-site variability for the small, fragmented marshes common to the Hunter Wetland National park. Although Sentinel-2 imagery has 13 spectral bands in the Visible Near Infrared (VNIR) and Shortwave infrared (SWIR) spectrum, the 10 m spatial resolution of Sentinel -2 is not enough to deal with the fragmented vegetation pattern in our study area. The earlier study (Rasel et al. 2016b) showed that high spatial resolution was important to deal with the spectral properties of saltmarsh species in a fragmented patch.

Another research area that is still challenging is to model biomass against remote sensing variables. Regression techniques are common to relate remotely sensed information (i.e. spectral bands or indices) with biophysical variables (i.e. biomass, leaf area index etc). However, traditional regression models have limitations to adequately capture the relationship and the spatiotemporal variability of the quantity (Kaheil et al. 2008) . Moreover, multicollinearity is an important issue for multiple regression model, especially when highly related variables (i.e bands of RS data) are selected as a predictor. Multicollinearity (also collinearity) is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, and Variance Inflation Factor (VIF) is used to detect multicollinearity before the process of multiple regression.

An ensemble method, random forest (RF), has reduced the problem of the multicollinearity issue (Liaw and Wiener 2002a) and has been proved to reduce of bias and overfitting (Breiman 2001) of a model. Considering two model approaches (bagging and boosting) of ensemble methods, RF works on both 'bagging' and 'boosting' algorithms (i.e.bootstrap aggregating) that aim to reduce the complexity of models that overfit the training data. Bagging techniques can be used to reduce the variance in model predictions, where numerous replicates of the original data set are created using random selection with replacement. Each derivative data set is then used to construct a new model and the models are gathered together into an ensemble. Clearly, the mean is more stable and there is less overfit. Therefore, the RF algorithm yields an ensemble that can achieve both low bias and low variance from averaging over a large ensemble of low-bias, high-variance but low correlation trees (Breiman 2001; Díaz-Uriarte and De Andres 2006b). RF originated from Classification and Regression tree (CART)(Breiman 2001) . However, decision trees suffer when features are correlated since they choose a single feature to maximize information gain at each step instead of a combination of features that also makes them unstable to small perturbations in the dataset. This instability is exploited in a very robust way in RF by building bagged tree ensembles. The multicollinearity problem

is alleviated since a random subset of features is chosen for each tree in a random forest (Cutler et al. 2007; Díaz-Uriarte and De Andres 2006a; Liaw 2002; Liaw and Wiener 2002b). This algorithm has been successfully employed in forestry for modelling quantities, forest biomass (Dube and Mutanga 2015) and wetland species biomass (Mutanga et al. 2012). Like RF, Support Vector Machine (SVM) is another machine learning approach that is already used in different fields of application (Brown et al. 2000; Kwiatkowska and Fargion 2003; Schölkopf and Smola 2002). For instance, Bruzzone and Melgani (2005) showed that SVMs reduced computational costs maintaining higher accuracy levels . A full description of both methods can be found in the following studies (Ben-Hur and Weston 2010; Breiman 2001; Camps-Valls et al. 2004; Smola and Schölkopf 2004).

Variable selection is a crucial issue in machine learning (RF and SVM) dealing with applied classification and regression problems(Hastie et al. 2001). There are three variable selection methods- filter, wrapper and the embedded method (Mehmood et al. 2012; Saeys et al. 2007). Detailed information of these three methods is available in Guyon and Elisseeff (2003) and Saeys et al. (2007). Very few studies (You et al., 2017) have been found where filter methods are used in combination with an RF classifier for land cover mapping. However, wrapper and embedded methods have been more commonly used with RF classifier (Dalponte et al. 2013; Genuer et al. 2010; Li et al. 2017; Millard and Richardson 2015). The RF itself provides three independent variable importance measures, Mean Decrease Accuracy (MDA) measure, the Gini Purity Index, and the number of times each variable is selected (Breiman 2001; Mansour et al. 2012). But the limitation of the RF algorithm, in measuring variables importance, is that it does not automatically select the optimal number of variables that produce the best classification accuracy (Adam et al., 2009). Moreover, Genuer et al. (2015) reported some classification problems when the variable importance index is based on the Gini Purity index of RF. In addition, Strobl et al. (2007a) argued that there is a bias in Random Forest variable measures in situations where potential predictors vary in their scale of measurement or their number of categories. This topic of variable selection still continues to be of interest. Addressing this issue, Díaz-Uriarte and De Andres (2006a) proposed an iterative backward feature elimination procedure to reduce the number of less relevant variables and Mansour et al. (2012) used a forward feature selection method embedded in RF for grass species degradation assessment from Hyperspectral data. The embedded approach Recursive Feature Elimination (RFE), has recently been used for feature importance in bamboo classification (Ghosh and Joshi 2014) from multispectral and landslides mapping (Stumpf and Kerle 2011) and from Hyperspectral data through the Random Forest classifier. Based on the documented limitations of RF, Genuer et al. (2015) proposed Variable Selection Using Random Forest (VSURF) packages in R to improve the feature selection accuracy. The embedded approach RFE has also been used with SVM for feature selection from Hyperspectral data (Pal 2006; Pal and Foody 2010) and known as RFE-SVM. RFE-SVM is a classical effective feature selection method by Guyon et al. (2002) that uses the coefficients in the SVM model to assess features, and recursively removes features that

have small criteria (Tang et al. 2016). It has both linear and nonlinear versions, according to the type of kernel function and commonly used algorithms for SVM feature selection.
To our knowledge, only a limited number of studies have been done, which explore these advanced feature selection methods in multispectral data. In contrast, there is ample proof for hyperspectral remote sensing for feature selection and dimensionality reduction (Pal 2006; Pal and Foody 2010). In this article, we present an embedded method, using recursive feature elimination (REF) for Random Forest and SVM classifier. However, variable selection for saltmarsh biomass estimates is challenging in terms of a single (i.e water) and multi-modal distribution (seasonal variation of vegetation) features of saltmarsh habitat. Therefore, we compared the performance of RFE with another feature selection method, VSURF to test the performance of these two methods for multispectral data, where there is still a research gap. The VSURF method has two advantages, (1) it finds out the important variables that are highly relevant to the response variables (Genuer et al. 2010); (2) it searches for a small number of variables sufficient for a good prediction of the response variables (Genuer et al. 2010). Byrd et al. (2014) also suggested wetland managers need to consider tradeoffs between cost, additional spectral information and the high spatial resolution of most commercial satellite imagery, which can identify within-site variability for the small, fragmented marshes common to the Hunter Wetland National park. Although many vegetation indices have been proposed in previous research (Dube and Mutanga 2015; Mutanga et al. 2012; Wang et al. 2016), depending on the diversity of species assemblage, indices vary in their relationships with biomass. However most of the important vegetation indices, like, Nitrogen Reflectance Indices (NRI), Green NDVI (GNDVI), Structure Insensitive Vegetation indices (SIPI), Normalized Difference Infrared Index (NDII) are similar to NDVI, which means they are also two band combinations. As we focused on the band importance of Worldview-2, we also calculated NDVIs involving all possible band combinations from the Worldview-2 imagery. In these circumstances, the primary objectives of this study were: to evaluate predictive models of aboveground biomass of a common saltmarsh plant species based on a low cost remote sensing data platform (objective-1); to evaluate different variable selection approaches to find out the best predictors for a response variable from multispectral data (objective-2); to use the spectral reflectance of Worldview-2 sensor as an input to compare two machine learning algorithms (Random Forest and Support Vector Machine), to estimate above ground biomass (objective-3).

7.2 Materials and Methods

7.2.1 Study site

The study area as shown in Figure 1, is located in Tomago, Australia (Longitude 151°43'40.6" E to 151°46'19.4" E and latitude 32°47'21.9" S to 32°51'29.4" S) which is approximately 8 km south of Raymond Terrace and 10 km north of Newcastle on the east

coast of Australia. The dominant tree species include *Avicennia marina* (Grey mangrove), *Casuarina glauca* (Swamp Oak or She-oak) and ground saltmarsh species includes *Sporobolus virginicus* (Salt couch) and *Phragmites australis* (Common reed). This study is only based on the training sample collected from the present distribution pattern (Figure 7-1) of salt coach (*Sporobolus virginicus*).

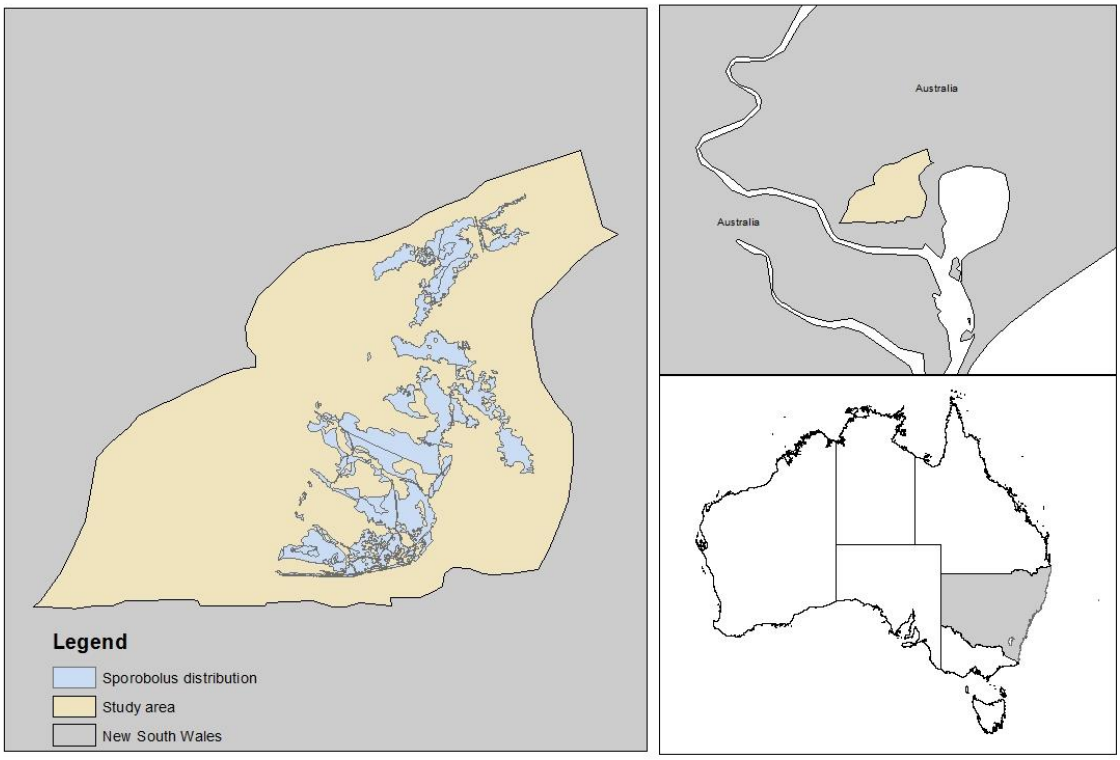

Figure 7-1: Australia country boundary and location of Wetlands (study site) in New South Wales (NSW) state. The first inset picture shows the Hunter Wetland National park in NSW (Clockwise), the second one shows the location of NSW state in Australia and the last and main view picture shows the study area with the distribution of Sporobolus virginicus species.

7.2.2. Field data collection

For Salt couch, a 20 m × 20 m vegetation plot was created to include a homogenous area of the same species. Three to five subplots (0.5m × 0.5m) were then randomly selected within each plot to measure the AGB. AGB was clipped within the subplots (0.5m × 0.5 m) and fresh biomass was then subsequently measured using a digital weighing scale. Based on these subplot measurements (n = 3–5), then average fresh above ground biomass per plot was calculated (Cho et al. 2007; Mutanga et al. 2012; Mutanga et al. 2004). The Worldview-2 satellite data used here were captured on 5th May 2015 that was the only cloud-free available data for this site. An extensive fieldwork campaign was conducted from 10th to 16th May 2015, just after the acquisition of the image. May is the dry month in eastern Australia and there was no surface water under the saltmarsh. Based on 0.5 % sampling intensity a total of 70 plots were required ((DoF 2010) for this study area. In total, 74 location (plots) were visited using a GPS. Each plot location was recorded with a meter level of accuracy of the global positioning system.

7.2.3. Remote sensing data acquisition and pre-processing

Worldview-2 satellite imagery was the primary remote sensing data used in this analysis. It has 0.46 m pixel resolution in the panchromatic mode and 1.84 m resolution in the multispectral mode. The multispectral mode consists of eight broad bands in the coastal blue (400-450 nm), blue (450–510 nm), green (510- 580 nm), yellow (585- 625 nm), red (630-690 nm), red edge (705 – 745 nm), NIR1 (770-895 nm) and NIR2 (860-1040 nm) parts of the electromagnetic spectrum. Digital number (DN) of Worldview-2 data has been converted to radiance data by applying the ENVI Worldview 2 calibration utility, available in ENVI v4.6 and later versions. It uses the factors from the Worldview-2 metadata and applies the appropriate gains and offsets in order to convert those values to apparent radiance. The FLAASH atmospheric module has been used in ENVI classic to remove atmospheric haze and to obtain surface reflectance data.

7.2.4. Extracting image spectra for model development

This paper adapted methods of extracting image spectra reported in earlier studies (Cho et al. 2007; Mutanga et al. 2012). Firstly, in situ data were collected in the study area. Then a window of 4x4 pixels (i.e 8mx8m) was used, as the vegetation plot, based on the GPS coordinates of each of the samples by overlapping the samples points and Worldview-2 images in ArcGIS 10.3. This window was used in ENVI Classic to extract the average spectra from each vegetation plot based on each spectral band of the Worldview-2 data.

7.2.5. Experiments for model calibration

At first 8 spectral bands from Worldview-2 data were selected as one of the inputs for the model and treated as Experiment-1. Then vegetation indices (VIs) were computed (Mutanga et al. 2012) from all possible two band combinations of Worldview-2. In this way total, (NxN=64) Vegetation Indices (VIs) were computed based on the following formula:

$$VI = \frac{(R_{(i,n)} - R_{(j,n)})}{(R_{(i,n)} + R_{(j,n)})} \qquad (7.1)$$

where R(*i,n*) and R(*j,n*) are the reflectances of any two bands from the selected bands for the spectral sample (n).

From these 64 VIs, 56 VIs were used as an experiment (Experiment-2). Here 8 VIs were removed from this Experiment considering their null value. Experiment-3 was the combination of experiments 1 and 2, so (56 VIs+8 bands of Worldview 2= 64) 64 variables were used as input. Then experiment-4 was based on the selected variables identified by the different variable selection methods from experiment 3 (Table 7-1).

Table 7-1: Four different sets used in this study.

Experimental Set	Data used	Description
1	Spectral bands	All bands of Worldview-2 data (8 bands, from coastal to NIR2 region)
2	Vegetation Indices (VIs)	All possible combinations (56) of VI from 8 bands of Worldview-2 data
3	VIs and Bands	Combination of 56 VIs and 8 spectral bands
4	Selected variables	Most important variables selected by variables section methods from the experiment 3.

7.2.6. Methodological flowchart for Random Forest and Support Vector Machine Regression

A general methodology summarized in figure 7-2, of the random forest (RF) and support vector machine (SVR) is described below:
(1) Data pre-processing: The spectral data were normalized to $0 \leq x \leq 1$ to avoid the over-weighting due to the features presenting the highest absolute values.
(2) Kernel: Different kernel types, linear, polynomial and radial basis function (RBF) were tested and the linear kernel was found to be most reliable to run the model. (This step is only for SVR).

(3) Cross-validation and grid search: Tune function of caret package was applied to tune *mtry* and *ntree* parameters for RF. Similarly, parameters C, γ and ε were solved by cross-validation and grid search on the training dataset. Then the value of epsilon (ε), gamma (γ) and cost (C) were determined to test the model. Details of these parameters will be discussed under the following subheadings of RF and SVM.

(4) Then the fine-tuned parameters were applied to the selected variables and performance (R^2 and RMSE) of the model were tested based on the validation data set.

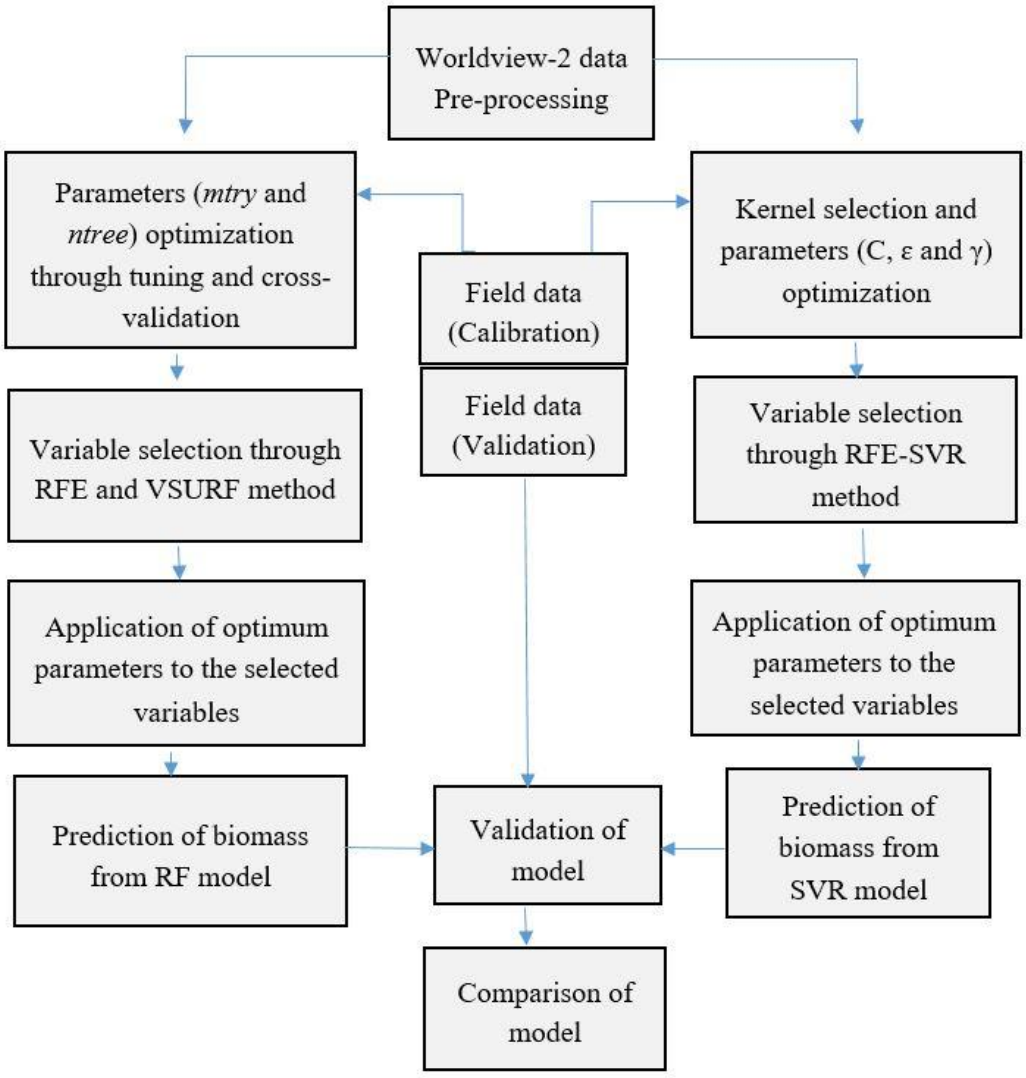

Figure 7-2: Workflow of machine learning algorithms for regression model.

7.2.6.1. Parameter optimization for the random forest (RF) and variable selection

RF works based on two tuning parameters, the number of trees in the ensemble (*ntree*), and the number of variables randomly sampled at each node to be considered for splitting (*mtry*) (Peters et al. 2002). In principle, one should simultaneously optimize both parameters before applying them to a model development. However, this computation process is intractable. We used the 'random Forest' library for RF regression, 'e1071' library for SVM regression, 'VSURF' (Genuer et al. 2015) and Classification and Regression Training (Caret) packages (Kuhn 2008) for feature selection. R statistical software (R development core team 2016) was used to tune the parameter, variable selection and execute the machine learning regression method. Two feature-selection methods used in this study are:

Recursive Feature Elimination (RFE): It is an optimization algorithm that aims to find the best performing feature subset from the original data set. At each iteration, it repeatedly creates models and keeps aside the best or the worst performing feature. Then it constructs the next model with the left features until all the features are exhausted. Finally, it ranks the features based on the order of their elimination. This method is used in RF and SVM for feature selection to resolve the problem of classification or regression.

Variable Selection Using Random Forest (VSURF): Genuer et al. (2010) used a two-step procedure for the VSURF method to select variables, where step -1 involves the preliminary elimination and ranking of variables and step-2 is the variable selection step through nested and predictive models. A brief explanation of these two steps is available in the Genuer et al. (2010).

7.2.6.2. Parameter optimization for Support Vector Regression (SVR) algorithm and variable selection

Epsilon SVR (ε-SVR): Standard SVR is the ε-SVR, where the sample points that support the "decision surface" or "hyperplane" are known as support vectors. These vectors fit the data according to the criteria of Epsilon (ε), gamma (γ) and cost (C) parameters. Here ε controlled the width of the epsilon-insensitive zone, used to fit the training data, and its value can affect the number of support vectors used to construct the regression function (Cherkassky et al. 1999; Vapnik and Kotz 1982). The value of ε determines the level of accuracy of the approximated function. It relies entirely on the target values in the training set. If epsilon is larger than the range of the target values, a good result cannot be expected. For example, if ε is 0, it will cause an overfitting problem. By contrast, the bigger the epsilon, the fewer support vectors are selected although bigger ε values result in more 'flat' estimates of the model (Durbha et al. 2007). The parameter cost (C) determines the balance between the model complexity (flatness) and the degree to which deviations larger

than epsilon (ε) are tolerated in the optimization formulation (Durbha et al. 2007). For example, if C is too large, then the objective is to minimize the empirical risk without regard to the model complexity part in the optimization.

Nu (v) Support vector regression: Schölkopf and Smola (2002) proposed a modification to the ε-SVR algorithm based on the difficulty in finding suitable values for the tube width ε. This modified method, called v-Support Vector Regression (v-SVR), automatically minimizes ε depending on the properties of the data (Axelsson et al. 2013). Here a new parameter (v) was introduced, in effect determining a fraction of the data points to be used as support vectors. The parameter C in the ordinary ε-SVR formulation is replaced by a parameter v which is bounded by 0 and 1. Earlier the parameter C could have taken any positive value, thus this additional bound is beneficial in implementation. The parameter v represents the lower and upper boundaries on the number of examples that are support vectors and that lie on the wrong side of the hyperplane, respectively.

RFE can be applied for SVR (Clevers et al. 2007) through a caret package based on the same principle of RFE and known as RFE-SVR. Beside RFE-SVR, Clevers et al. (2007) proposed another method named SVM band shaving that is also based on the RFE approach of classification problem (Guyon et al. 2002). In this study, we used RFE-SVR based on Caret package to optimize SVM parameters.

7.2.7 Model validation and accuracy assessment

K-fold cross-validation is implemented in this study to optimize the parameters for both algorithms (RF and SVR), based on training data sets (70% of the whole data). Finally, the calibrated model that comes from k-fold cross validation is tested once again with the independent test dataset (30% of the data) to estimate RMSE, R^2 and bias of the model.

7.3 Results

7.3.1 Descriptive statistics of biomass (kg/m^2)

Standing biomass of Salt couch ranged from 0.450 kg/m^2 to 2.175 kg/m^2. The average biomass was 1.3828 kg/m^2 with a standard deviation of 0.3899 kg/m^2. Based on the Shapiro-Wilk normality test (W= 0.9818, p value = 0.3673) and skewness (-0.2499402) and kurotosis (-0.2130938), it was found that biomass (kg/m^2) was normally distributed (figure 7-3) .

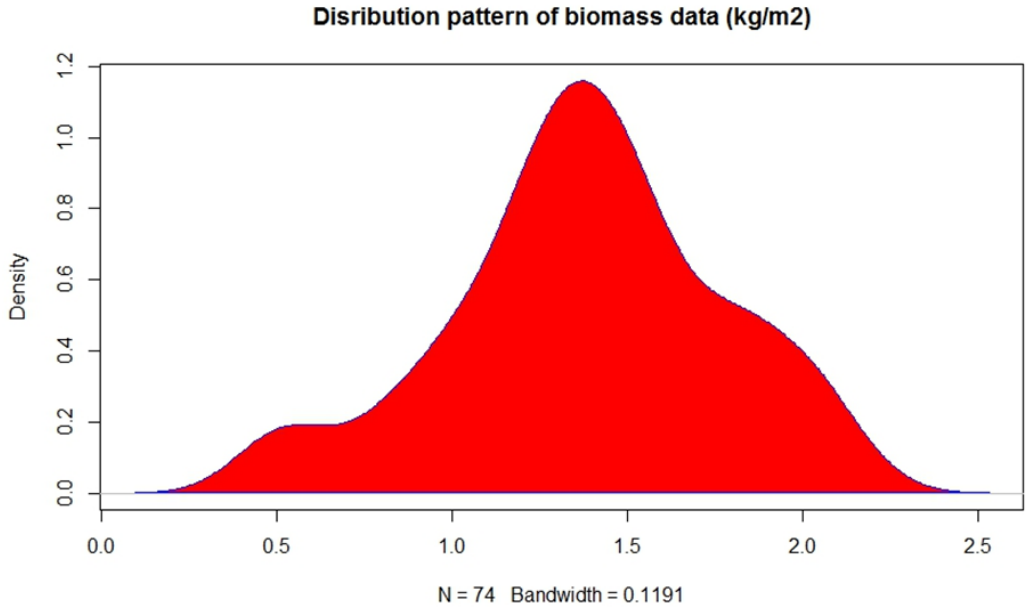

Figure 7-3: Biomass (kg/m2) of Sporobolus virginicus distribution pattern shows that data are normally distributed.

7.3.2 Optimum parameter of random forest (RF) algorithm to estimate biomass

Instead of default (1/3 of the total number of variables), *mtry*, lowest Out-Of-Bag (OOB) error rate was used to determine the best value of *mtry* (table 7-2). In the OOB method, some of the training data are excluded for each regression tree generation, and the errors for these data can be used to inform the RF of the relative strength (Dube and Mutanga 2015).

Table 7-2: Maximum and minimum OOB errors obtained using different mtry values.

mtry	3 (minimum)	21 (default)	28 (optimum)
ntree (default)	500	500	500
OOB error rate	42.64	33.23	32.13
Variability explained (%)	57.36	66.77	67.87

Here cross-validation in parameter optimization was done to select the best mtry value based on RMSE. Best on the cross-validation result (figure 7-4), mtry value 28 was selected as the best on the lowest RMSE (0.227 kg/m^2).

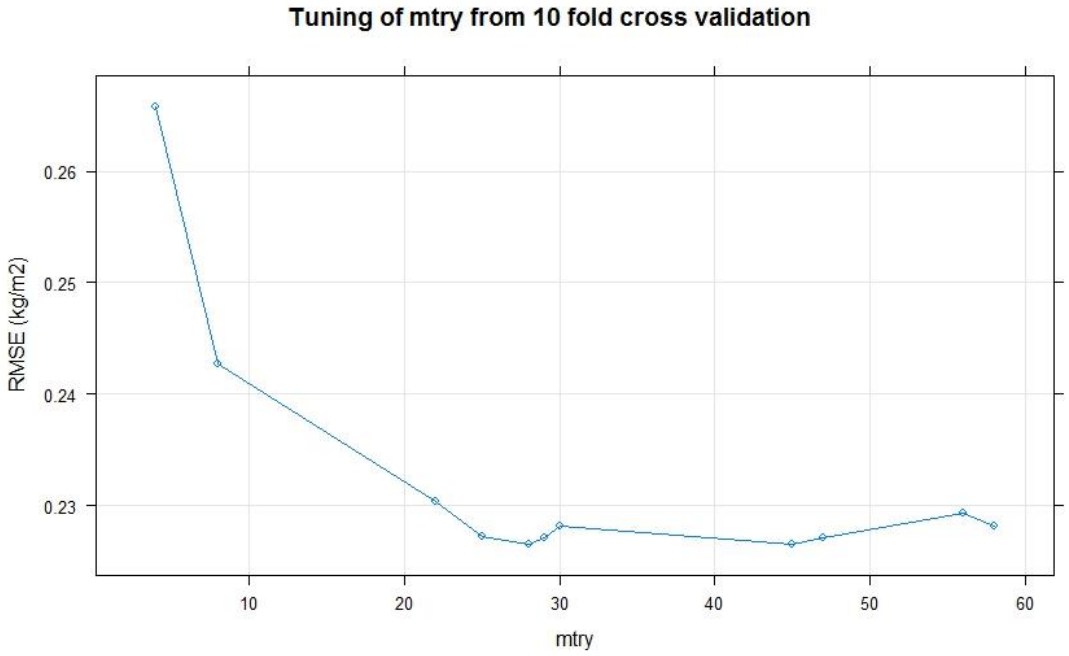

Figure 7-4: Optimization of random forest parameters (mtry) based on RMSE (kg/m2).

The earlier *mtry* parameter was tuned as it has an effect on the final accuracy. Whereas the *ntree* parameter is different, in that it can be as large as the user wants, and continues to increases the accuracy up to a certain point. From the figure 7-5, it is clear that *ntree* have a very minimal effect on the results of RMSE (kg/m^2) and R^2. We can see that the most accurate value for *ntree* was perhaps 4000 with a mean RMSE of 0.2270 kg/m^2. It held a constant value for *mtry* (i.e 28) that was optimized in the previous step. Based on the cross-validation result, *mtry* value =28 and *ntree* = 4000 are acceptable as both value provided the same RMSE (0.227 kg/m^2).

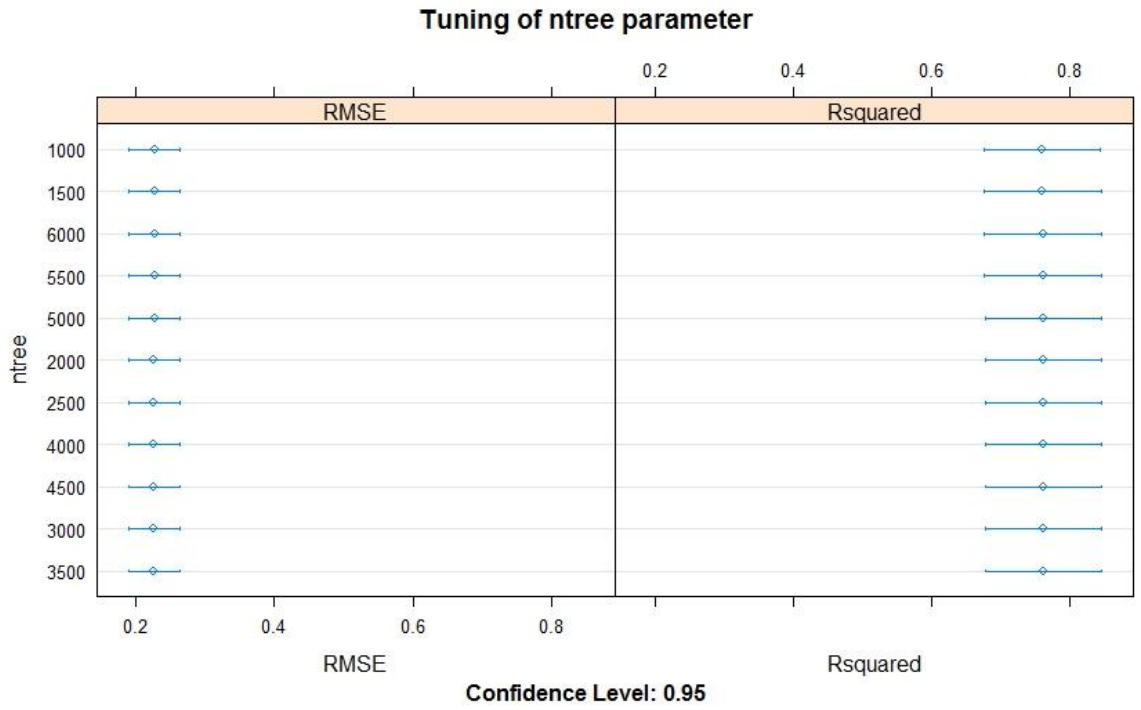

Figure 7-5: Tuning of the Random Forest (RF) parameter (ntree) for Sporobolus virginicus biomass (kg/m2) estimation.

7.3.3 Variable Selection From Different Methods in Random Forest

7.3.3.1 Recursive Feature Elimination (RFE) method

The recursive feature elimination (RFE) was able to identify the smallest number of explanatory variables that would offer the best predictive ability of the random forest (figure 7-6). Here four variables were selected and those are (a) Near Infrared II (Band 8), (b) Near Infrared I (Band 7), (c) Red-edge (Band 6), and (d) Yellow (Band 3). The use of these four variables produced the lowest RMSE using 10-fold cross validation (0. 166 kg/m^2), while the use of entire variables (n = 64) produced the highest RMSE for 10-fold cross validation (0.211 kg/m^2) (Figure 7- 6).

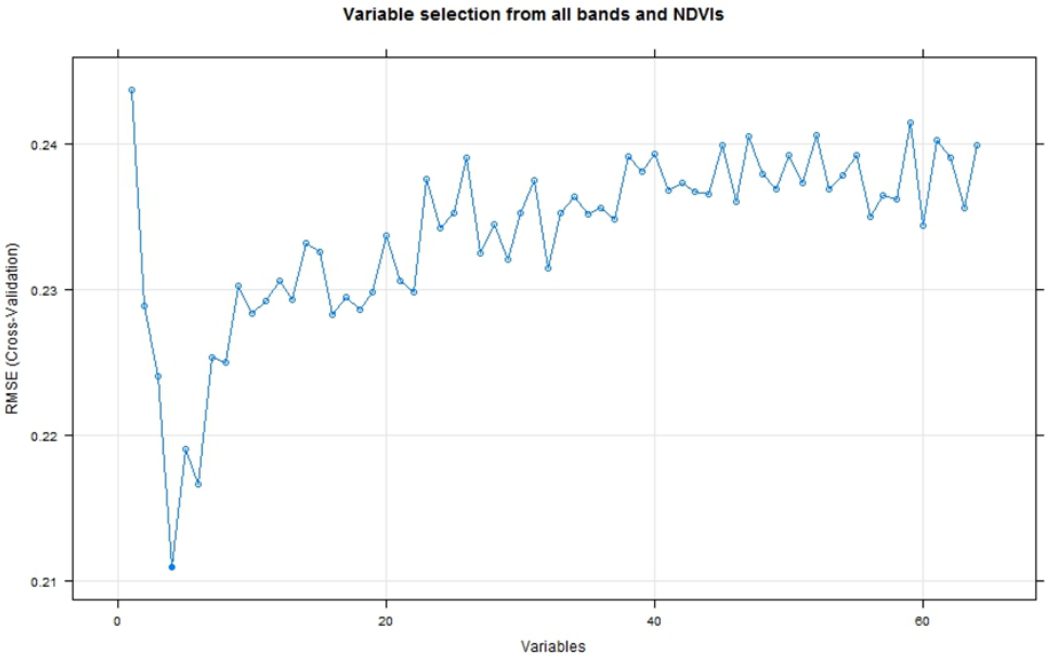

Figure 7- 6:The RMSE obtained during the recursive variable selection (RFE) process. The lowest RMSE (kg/m^2) obtained from the lowest number of variables (here 4 variables are selected).

3.3.2. *Variable Selection Using Random Forest (VSURF) method*

In the VSURF method, the important variables (1 to 10) are respectively selected (figure 7-7) based on the optimized model parameter of *ntree* = 4000 and *mtry* = 28, which are values of the main RF model previously determined. Four sequential plots of Figure 7-8 describe the four main steps of the VSURF algorithm for variables selection (Genuer et al. 2015). Four steps of figure 6 are discussed below:

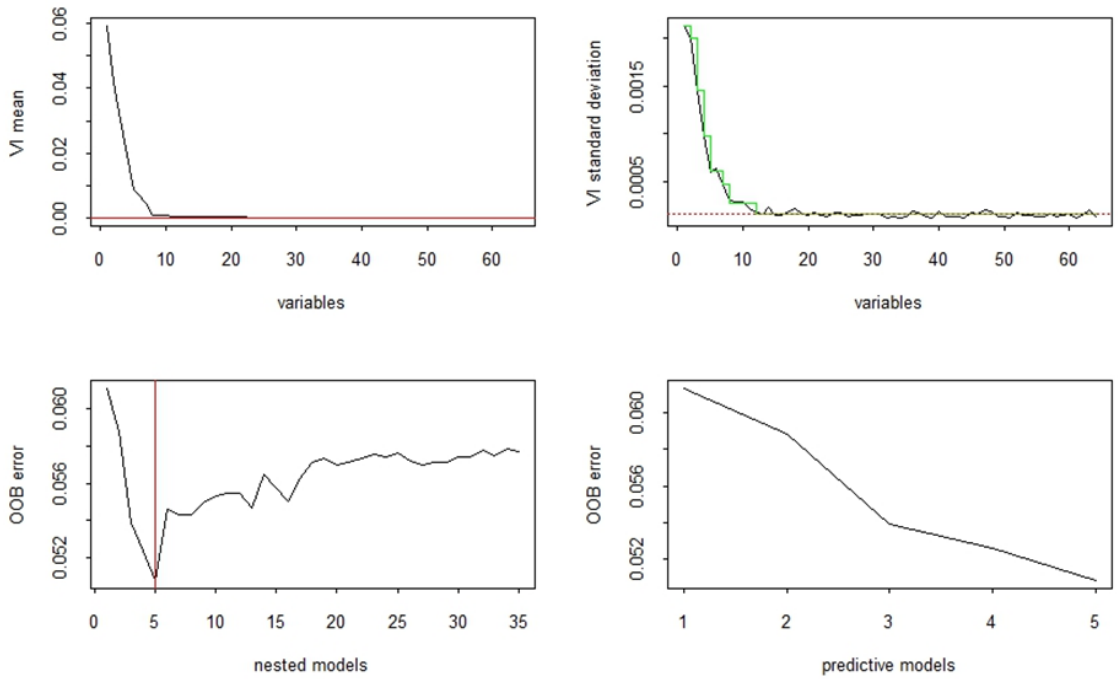

Figure 7-7: Variable selection procedures for interpretation and prediction for Sporobolus virginicus biomass (kg/m^2) data based on VSURF function.

Variable ranking (Figure 7-7, top left): Here the variables were ranked by sorting the Variable Importance (VI) in descending order. The results (presented on the top left graph) for all variables (n=64) shows that true variable are significantly more important than the noisy variables.
Variable elimination (Figure 7-7, top right): The standard deviations of variable importance (VI) can be found in the top right graph of figure 6. It is clear that true variables standard deviation is large compared to the noisy variables. This step retains more variables (here 20 variables) than necessary in order to make a careful choice later (Genuer et al. 2010).
Variable selection procedure for interpretation (Figure 7-7, bottom left): This graph shows the OOB error rate of RF (using default parameters of RF) of the nested models starting from the one with the single most important variable, and ending with the one involving all important variables kept previously. Ultimately, the selected variables of nested models lead to the smallest OOB rate. It is clear that the error decreases quickly and reaches its minimum when the first 5 true variables are selected in the model.

Variable selection for interpretation (Figure 7-7, bottom right): The graph shows the results of variable selection for the prediction model. In the predictive model, a variable is added only if the error gain exceeds a threshold. Here RMSE (kg/m^2) starting from 0.245 (MSE 0.060) decreases up to 0.228 (MSE 0.052) when five variables added to the model. This RMSE is slightly higher than the RMSE of the RFE method.

7.3.4 Optimum parameter of Support Vector Machine (SVM) algorithm and variable selection

Tune function shows that cost = 5.0, epsilon = 0.2 and gamma= 0.5 were found most suitable (figure 7-8) for SVM parameter. Where dark blue indicates lower RMSE and lighter blue indicates higher RMSE. Figure 9 shows the effect of a decreasing number of variables in terms of RMSE (kg/m^2) for fresh biomass. From this figure, we can conclude that the predictive power of a model with the least number (here four) has a RMSE (kg/m^2) less than 0.180 kg/m^2, which can be considered as good for this dataset. These four variables were the following bands: (a) Near Infrared II (Band 8) (b) Near Infrared I (Band 7) (c) Red-edge (Band 6) and (d) Yellow (Band 3).

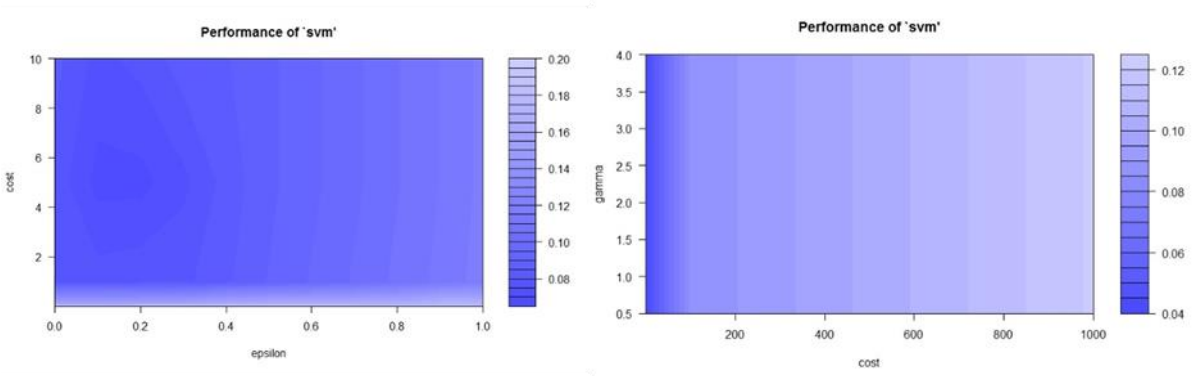

Figure 7-8: Optimization of SVR parameters (cost, epsilon and gamma) based on RMSE (kg/m^2).

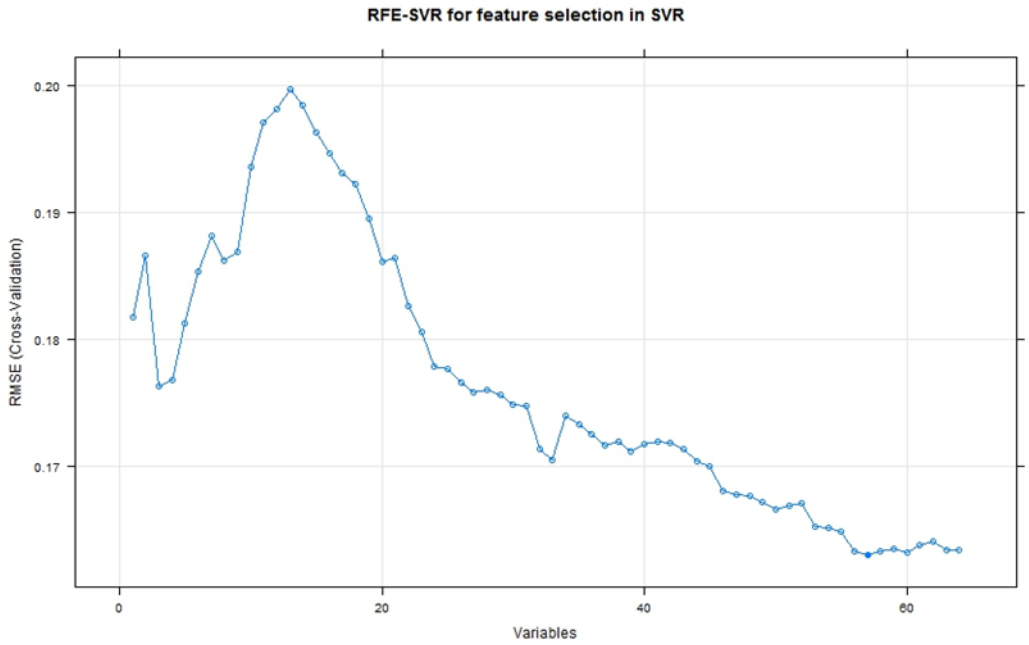

Figure 7-9. RMSE (kg/m^2) results of the RFE in SVR for biomass with a varying number of features (all bands and all vegetation indices).

In the feature selection, RFE- SVR was developed based on a linear kernel (Table 7-3), even though it provided good results for feature selection in terms of RMSE and least number of variables. This method provided the lowest RMSE (0.1761 kg/m^2) using the three variables. When a 4th variable was added, there were a very minor change in RMSE (0.1777 kg/m^2). Therefore, to maintain the consistency of variable selection between RF and SVM, we also picked 4 variables from the RFE-SVR method.

Table 7-3: Results of the biomass prediction in terms of RMSE (kg/m^2) using different methods based on an entire variable (n=64) of calibration data.

Method of feature selection	No of selected variables	RMSE (Kg/m^2)
RF (Default)	5	0.227
RFE	4	0.211
VSURF	5	0.228
RFE-SVR	4	0.176

7.3.5 Performance of Machine Learning Regression (MLR)

The results in Table 7-4 shows above ground biomass estimates for Salt couch derived from the spectral band (n= 8) and vegetation indices (n=56) of Worldview-2. It is clear that high AGB estimates were obtained from the use of the selected spectral bands (Experiment set iv), compared to the other three experiments where different vegetation indices and combinations of vegetation

indices and spectral bands were used. It can be observed that the random forest regression produced the highest R^2 (0.72) and the lowest RMSE (0.166 kg/m^2) using the four important predictors, (a) Near Infrared II (Band 8) (b) Near Infrared I (Band 7) (c) Red-edge (Band 6) and (d) Yellow (Band 3). Although an optimized parameter (Figure 7-7) of ε-SVR with linear kernel worked well for calibration data, it provided unreliable results when treated with a validation dataset using other different kernel types. Hence, based on the recommendation of Axelsson et al. (2013), v regression was found more reliable for biomass prediction. Similarly, SVM also produces the highest R^2 (0.66) and lowest RMSE (0.200 kg/m^2) from the experiment set-iv with the four selected bands (table 4), whereas the lowest performance resulted from the experiment set ii where only vegetation indices (n= 56) were used.

Table 7-4: Summary of above ground biomass (ABG) estimations from two machine learning approaches based on four different experiment sets.

Experiment Set	Validation set (n=22)		
RF regression model	R^2	RMSE (kg/m^2)	Bias
I (Only spectral bands)	0.70	0.202	-0.03859945
II (Only VIs)	0.19	0.294	0.102
III (All bands and VIs)	0.67	0.188	-0. 0237859
IV (Selected variables)	0.72	0.166	-0.0559891
Nu SVR regression			
SVM	R^2	RMSE (kg/m^2)	Bias
I (Only spectral bands)	0.66	0.208	-0.0480
II (Only VIs)	0.008	3.270	-3.2579
III (All bands and VIs)	0.63	0.236	0.08041
IV (Selected variables)	0.66	0.200	0.03553

A one to one relationship between actual and predicted biomass, for both regression models, is shown in Figures 10 and 11 for RF and SVR respectively. The RF model, based on all four experiments, tended to underestimate the high biomass values that fall beyond the range (figure 7-10). Whereas in all four experiments the SVR model overestimated that value (figure 7-11).

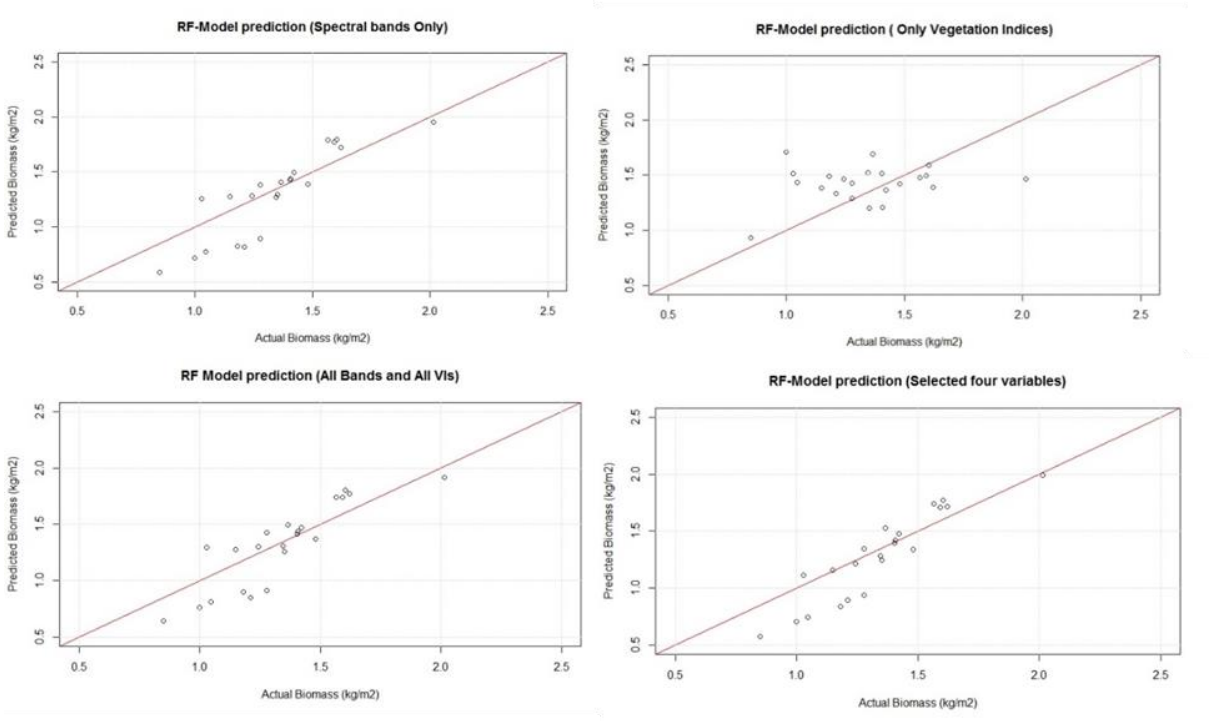

Figure 7-10: Relationships between actual and predicted biomass of Sporobolus virginicus using the random forest regression model.

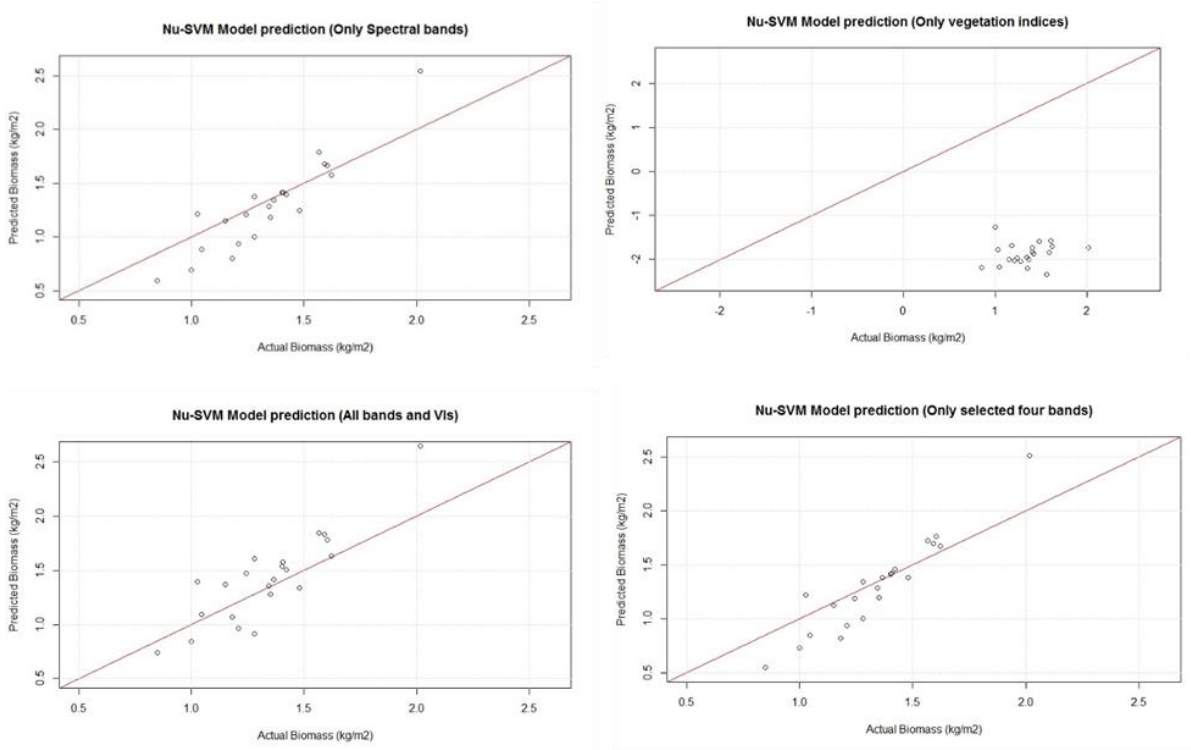

Figure 7-11: Relationships between actual and predicted biomass of Sporobolus virginicus using the v-SVR regression model.

7.3.6 Comparison of RF and SVR model

Descriptive statistics developed from two machine learning models are presented in table 5 to compare with the validation test data collected from the field. The minimum standard deviation was observed in the RF model.

Table 7-5: Comparison of descriptive statistics derived from field observation and model-derived data.

Model	N	Max	Min	mean	SD
Field data	22	2.17	0.4505	1.38	0.390
RF	22	1.98	0.5914	1.276	0.383
SVM	22	2.51	0.5487	1.302	0.427

A scatterplot matrix (Figure 712) was developed when results of RF were compared to SVR results, based on four selected variables only. This matrix is valuable when considering whether the predictions from two different algorithms are uncorrelated. If weakly correlated, they are good candidates for being combined in an ensemble prediction. For example, RF and SVR look weakly correlated.

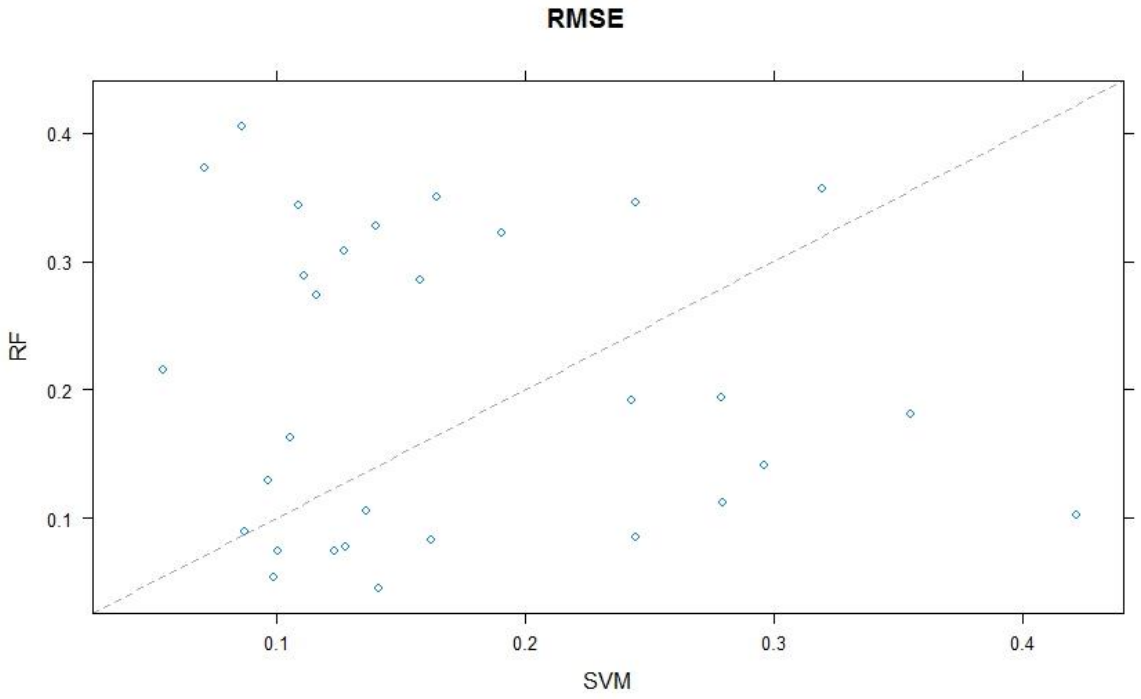

Figure 7-12: Comparison of Machine Learning Algorithms (SVM and RF) in scattering Plots.

We can see pair-wise statistical significance scores (Table 7-6). The lower row of the table shows p-values for the null hypothesis (distributions are the same), where smaller is better. When the value of the upper row is 1.0, there is no difference between the models. The lower row indicates the p values from the significance test. We can see a significant difference in RMSE (kg/m^2) between SVM and RF. We can also see little difference between the distributions of R^2 for SVM and RF, although they are not statistically significant at the 95% confidence interval.

Table 7-6: Statistical differences between two models

Model	RMSE		R^2	
	SVM	RF	SVM	RF
SVM		-0.03427		0.07336
RF	0.000022		0.0006525	
Upper row: estimates of the difference; Lower row: p-value for H_0: difference = 0				

7.4 Discussion

7.4.1 Performance of Worldview 2 spectral bands for biomass estimation

Based on the four experiments, selected variable (four selected bands) based machine learning regression clearly produced a more robust fit than different vegetation based indices based regression. Results from this study indicate that the Salt couch species biomass is accurately estimated with the spectral information of the new red-edge (Band 6), NIRI (Band 7) and NIRII (Band 8) bands. In this commercial multispectral data, the red-edge is one of the additional spectral bands that detects energy on a narrow band between 705 and 745 nm at 1.84 m spatial resolution. Similarly, in a different study, Eckert (2012) showed that the texture measures mean derived from band 3(green), band 4 (yellow), band 6 (red edge), and bands 7 and 8 (both NIR bands) indicate a strong relationship with biomass and carbon. Considering the benefit of the additional 4 bands of Worldview-2, Digital Globe (2010) mentioned that although NIR II (860 -1040 NM) overlaps with the band NIR I (770- 895 nm), NIR II enables biomass analysis as it is less affected by the atmospheric influence. Wolf (2010) also mentioned that the red band stays true to lower reflectance levels than the NIRII band, which has a higher value than traditional broad NIR bands, hence a higher NDVI value is produced that is important for any vegetation property analysis. That is why both NIRI and NIRII bands were useful to improve the accuracy of biomass estimation. Marshall et al. (2012) suggested that the yellow band of Worldview-2 would be very important to discriminate Buffalo grass if the image had been acquired in dry the season. In our study, the Worldview-2 image was acquired at the onset of the dry season and played a significant role in estimating biomass.

Our results are supported by Lawrence and Ripple (1998) and Gadallah and Csillag (2002) when they regressed band wise information against above ground biomass using aerial photo and Landsat TM respectively. Recently Dube and Mutanga (2015) found a better R^2 (0.33) and RMSE (63.61%) from a combination of spectral information and vegetation indices rather than the individual use of spectral information (R^2 = 0.27, RMSE = 67.15%) and VIs (R^2= 0.23, 68.28%) when they estimated forest biomass from the latest Landsat 8 OLI data. One possible explanation is that the band wise regression approach allows for the decoupling of bands and make the analysis easier to discover different relationships between the response variable and each band, including different polynomials, coefficients, and transformations (Lawrence and Ripple 1998). However, this flexibility is not possible with regression against vegetation indices (Lawrence and Ripple 1998).

There are other findings where different authors used different vegetation indices to predict saltmarsh biomass. For example, (Gross et al. 1986) and (Hardisky et al. 1983; Hardisky et al. 1984) found spectral radiances based indices worked well to predict biomass for *Spartina alternflora*, one of the tallest saltmarsh varieties. Because tall vegetation structure increases light scattering and absorption in spaces between the vegetation(Byrd et al. 2014) leading to lower overall canopy reflectance (Mutanga and

Skidmore 2004a). In a recent study, Mutanga et al. (2012) showed that different modified VIs can predict biomass of wetland, although they did not compare those VIs with spectral information separately. However, their findings showed that instead of traditional NDVI, modified VIs (three VIs generated from band 6 and 7) worked better to predict wetland biomass. VIs that are associated with band 6 (Red-edge) and band 7 (Near infrared band) resulted in robust predictions for a high-density wetland biomass (Mutanga et al. 2012) dominated by *Cyperus papyrus* and *Phragmites australis*, two tall varieties. Our findings also revealed that band 6 and 7 were two important variables to predict biomass of *S. virginicus*, a spreading perennial that is rarely affected by the obstacles associated with taller varieties. This may be another reason why spectral bands based machine learning performed better, in comparison to VIs and combination of VIs with spectral bands.

Our findings are also supported by Mutanga and Skidmore (2004b), where they indicated that the red edge (700–750 nm) and longer wavelengths of the red edge (750–780 nm), yielded higher correlations with biomass ($R^2 = 0.77$) than the standard NDVI. In a different study, Cho et al. (2007) who found the better predictive performance of the red-edge extracted from airborne HyMap imagery in estimating grass/herb biomass in the Majella National Park. Our results indicate how tremendous saving of money can be made as well as reductions in time spent on hyperspectral data acquisition and processing. Moreover, the results from this study also confirmed the tradeoffs between cost, additional spectral information and the high spatial resolution of most commercial satellite imagery that was one of the recommendations of Byrd et al. (2014). Clearly, it can be questioned whether similar results can be found for more complex and mixed species ecosystems, but this can be evaluated in future studies.

7.4.2 Performance of variable selection methods and machine learning algorithms

Among the three different algorithms of variable selection, it was found that recursive feature elimination provided the lowest RMSE (kg/m^2) with the least number of variables. Although RFE-SVM selected three variables and we added the 4^{th} variable to keep the consistency with RFE-RF, there was no change in the result when RFE-SVM was run with the three variables. This was the limitation of RFE-SVM and this was also supported by Tang et al. (2016). When some of the candidate features are highly correlated, the assessing criteria of the importance of these features will be underestimated (Tang et al. 2016) and this limitation is called "correlation bias". That was the main cause of poor performance of RFE-SVM compared to RFE-RF. The best performance of RFE is also demonstrated by (Ismail et al. 2010) when they found an OOB error rate of 8.34% from 14 variables in comparison with OOB error rate of 8.23% from 21 variables using a backward selection approach.

From the findings of our study, it is clear that RF performs better than SVM. As these differences are statistically significant for RMSE (kg/m^2), there are a couple of arguments for why RF performed better than SVR. It might be due to the degree of multicollinearity, and how well the SVR parameters are tuned (Axelsson et al. 2013). As mentioned previously, RF deals with small subsets to find the pure variables, omitting the multicollinearity problem of regression analysis, and provide better predictions. Two band combination vegetation indices used in this study are correlated, however, as

demonstrated by Cutler et al. (2007), RF is not sensitive to collinearity. Moreover, the grid search used for parameters tuning was automatic and the tuning process of ε-SVR and v-SVR proved to be difficult and time-consuming (Axelsson et al. 2013). It took a long time to learn how to set the intervals and boundaries of the grid search properly in R interface. In addition, the optimized parameters of ε-SVR were providing overestimated predictions for biomass, and alternatively, v-regression was selected based on the recommendation of (Axelsson et al. 2013).

There are other studies (Axelsson et al. 2013; Clevers et al. 2007) where authors revealed the better performance of SVR when it was compared with Partial Least Square (PLS) based on transformed (i.e continuum removal) data. But in contrary, Marabel and Alvarez-Taboada (2013), showed that SVR outperformed PLS regarding RMSE when no transformations were applied to reflectance data. However, in this research, we did not do any transformation of reflectance data and it was not compared with PLS. One aspect of SVR is essential to consider that the accuracy of the SVR model was influenced not only by its parameters but also by the spectral region used as variables (Marabel and Alvarez-Taboada 2013). This argument was confirmed in our finding when fine-tuned parameters of calibration model failed to provide a reliable estimation for validation data.

Although we compare, for the first time, use of RF and SVM regression for remote sensing of saltmarsh biomass estimation, further investigations are required to improve prediction performance of the model. In our results, the random forest tended to underestimate the high biomass values that fall beyond the range whereas SVM overestimate that value that might be a subject for further studies using additional data sets. Results demonstrated that SVR and RF regressions, are both used to predict biomass. However, the ability of automatically producing accuracy assessments and measuring the variable importance make Random Forest algorithms more effective.

7.5 Conclusions

Our results have shown that the commercial multispectral sensor Worldview-2 is a potential data-source for wetland managers for an acceptable (R^2 = 0.72, RMSE = 0.166 kg/m^2) saltmarsh biomass estimation model. Recursive Feature Elimination (RFE) is a more powerful, robust and dependable technique that can select variable for model prediction. Both of the machine learning algorithms (RF and SVR) perform well to predict biomass for saltmarsh species. However, RF outperformed SVR in respect of RMSE due to its special settings of the subset that finds pure variables omitting the multicollinearity problem. Although there are some recent findings of Worldview 2 for wetland biomass estimation, this result was developed based on common saltmarsh focusing on a species-specific model. Moreover, two famous machine learning approaches were compared to make them user-friendly at the local level (i.e. wetland manager). Although further investigation is recommended for ε-SVR with additional data to check its behaviour with independent data set, v-SVR was an alternate option to estimate biomass. The results of this work will be helpful for reliable aboveground biomass estimation for wetland managers and ecologists.

Chapter 8

Synthesis: High spatial resolution multispectral image and machine learning algorithms for saltmarsh classification and biomass modelling

This chapter is the synthesis of the whole thesis. It also highlights the special contribution of the author in this research.

Up to mid-nineteenth century, saltmarsh ecosystem was treated as wastelands and boggy swamps. Many saltmarsh areas have drained and lost due to a lack of recognition of its ecological value and less attention to their restoration (Gedan et al. 2009). It has attracted more attention as the understanding of its ecological value has improved over the past few decades and now treated as significant ecological communities that provide key habitat for other marine fauna (Adam 1993). Therefore, monitoring and mapping of wetland plant species distribution are important in wetland management, conservation and restoration . This rigorous task requires intensive field-work, a collection of ancillary data, and the visual estimation of percentage cover for each species to be classified. This is time-consuming, costly and labour intensive (Lee and Lunetta 1995). To overcome these limitations, remote sensing offers a practical and economical means of plant species classification and estimates the biochemical and biophysical parameters of the wetland species. Remote sensing helps improve the efficiency and effectiveness of fieldwork for monitoring the temporal changes of wetland habitats.

Selecting suitable remote sensing systems and their data are important for mapping land use and land cover (LULC). The accuracy of classifying saltmarshes and other wetland plant species varies based on the spatial and /or spectral resolutions of the remotely sensed data. Mostly pixel-based classification method of multispectral image dominated by mixed pixel are often incapable of producing accurate species classification. Traditional pixel-based classification methods assume that the reflectance of each subclass is normally distributed. This is a common violation in remote-sensing data, especially when classes or even subclasses of the main class contain different spectral features (Kavzoglu and Reis 2008). To overcome this problem, non-parametric classifiers such as decision tree classifier (DTC), artificial neural networks (ANN), and Support Vector Machine (SVM) are getting more attention in recent classification techniques applied to remotely sensed data (Zhu and Blumberg 2002; Kavzoglu and Reis 2008; Otukei and Blaschke 2010). Especially SVMs are capable to handle small sample size data for training. But a large number of parameters settings and kernel selection are the major limitations of SVMs. Alternatively, random forest (RF) is a relatively new ensemble classifier that has a minimal number (only two) of parameters and embedded variable selection method that are easy to apply. The variable selection is another important part of the machine learning algorithm and their applications. Based on the literature review of this thesis, only a very few research has been done where advance variable selection methods were applied for saltmarsh classification and biomass estimation. Therefore, this thesis intends to fill this research gap.

As a part of a suitable sensor selection, the author compared spaceborne hyperspectral data (EO-1 Hyperion) with multispectral Worldview-2 in chapter 2. Four different images of a spaceborne hyperspectral data (EO-1 Hyperion) scenes from around the coastal region of Australia were tested, it shows that there is a strong relationship between the acquisition time of year and the SNR of the Hyperion data. Calculated SNR for Hyperion SWIR data is higher in the summer and lowest in the winter (Figure 2.5 and 2.6) that was similar to the finding of Kruse et al. (2003). Chapter 2 of this thesis explored the maximum number of

endmembers that can be retrieved from the Hyperion data. Based on the Principal Component Analysis (PCA) of atmospheric effect corrected reflectance data, the first 16 principal components were found that contained most (99.83 %) of the information. It was similar to the findings of Chauhan et al. (2011) and Pervez and Khan (2015). In this study, only 56 bands from the Very Near Infrared (VNIR) bands that were similar with the wavelength of Worldview-2 data were selected for PCA. Among the 56 bands of atmospherically corrected Hyperion data, 79.01% variability was contained by the first PC, 96.31% by PCs 1-2 and likewise 99.83% by PCs 1-16. PCA highlights the redundancy in data due to similar responses in some wavelengths and reduces the dimensionality of data. When 16 selected bands of Hyperion data were compared with the 8 bands of Worldview-2 for saltmarsh classification, the overall classification accuracy has increased in both cases after adding band orderly. But the overall accuracy obtained from Worldview-2 was higher than that from the EO-1 Hyperion image. Table 2-6 shows that OA % for Worldview-2 was increased from 72 to 79 while for Hyperion it increased from 70.47 to 71.66. Considering the significance test with z values and kappa statistics at 95% confidence interval, Worldview-2, classification accuracy was higher than Hyperion data. So this study explored that the importance of spatial or spectral resolution is depends on the feature of the study area. Although hyperspectral data have narrow bands to retrieve a full spectral profile, due to small patch size of vegetation pattern, 30-meter spatial resolution was not effective to classify them using EO-1 Hyperion data. As the spatial resolution was the important consideration for saltmarsh classification, the author considered two sampling methods, pixel-based and region of interest (ROI) from the Worldview-2 and Landsat-8 OLI data discussed in chapter 3.

The distribution of classes in feature space is an important factor related to the classification accuracy. Because class probabilities are produced based on the distribution of classes in feature space. However, different classes often have overlaps which reduce the accuracy of classification. To address this problem, the concept of spectral separability is introduced. A very few research studies (Collin and Planes 2011) worked on spectral separability analysis for saltmarshes species community. Selection of spatial unit (pixel) and collection of training samples also affect the separation between two species classes within a feature space and have an impact on the accuracy of the thematic map. determine how much distance are present within a feature space for two species and regulate the thematic accuracy of the map. Therefore, chapter 3 investigates the spectral separability derived from the region of interest (a group of homogenous pixels) or individual pixel based training sample. In addition, it also explores the potentiality moderate resolution (Landsat 8 OLI) broadband data for two dominant species, Grey Mangrove (*Avicennia marina*) and She-Oak (*Casuarina glauca*).

Based on spectral curve originated from the region of interest (ROI) sampling, it was evident that all saltmarsh species were overlapped within all bands of the spectrum for Landsat 8 OLI data. But Mangrove and She-oak (*Casuarina glauca*) tree species were clearly separable from each other up to bands 5. Because Mangrove and She-oak are two dominant tree species cover a large extent of the study area that is easily detectable by 30-meter pixels of Landsat data. On the other hand for Worldview-2, perennial grass (GR) was clearly separable in green, red edge and NIR bands. Reflectance properties of water (WA) in the last 3 bands (Red Edge, NIR1 and NIR 2) and reflectance of wetland soil (WS)

in band 8 are significantly different from other classes. Two saltmarsh species (*Phragmites australis* and *Sporobolus virginicus*) were clearly separable in the red band when pixel based sampling was followed for separability analysis. The saltmarsh species also show band-specific within-species variance. For example, the variance of *Phragmites australis* in the coastal band is quite large and in NIR2 is relatively small. Based on the spatial unit (pixel) distribution, it is clear that there are no outliers for any species in any band (figure 3-4). Moreover, the median is the centre of each box indicating that data are normally distributed. Further research can be carried out to evaluate the performance of different classifiers and machine learning algorithms. Because most of the time, saltmarsh spatial distribution is multimodal and need advance classifier to handle this distribution pattern. To address this issue the author reviewed most of the works of machine learning classifiers related to saltmarsh and wetland classification and was discussed in chapter 4.

Chapter 4 reviewed the application of machine learning and advance classifiers for saltmarshes and wetland ecosystem. Maximum likelihood (MLC) is the most common supervised classification method. However, in most of the cases, the normal distribution of the classes in the spectral domain is the main violation of remotely sensed data. Because sometimes the reflectance values of the main class and their subclass contain different spectral properties that make the application of MLC more difficult and less accurate. In addition, adequate ground truth information and collection of sufficient samples of training and validation are also impractical for wetland ecosystem. Therefore, the application of machine learning algorithms is increasingly popular for land cover mapping. Although SVM related remote sensing studies were reviewed by Mountrakis et al. (2011), other types of machine learning methods have been advanced since then. New articles have been published since then, and new methods have been used for SVM. Very recently, Belgiu and Drăguţ (2016) reviewed Random Forest and its application in remote sensing. But the potentiality of ensemble classifiers for a specific ecosystem like saltmarsh and wetland ecosystem need to be reviewed. That is why, emergence and modification of machine learning algorithms and techniques in recent years necessitate such a review, which will be highly valuable for guiding or selecting a suitable classification procedure for a specific ecosystem.

Among the machine learning algorithms, SVM is appealing due to its impressive ability to successfully handle a small training dataset while producing higher classification accuracy. It is proposed as a superior classifier in remote sensing than the traditional methods like Maximum Likelihood Classifier (MLC) (Montero et al., 2005). Although there is ample proof of training sample size reduction for SVM, very few research are available for wetland ecosystem where data collection is really a challenge due to hazardous access. In addition, kernel selection for SVM and its relation with spectral separability for each species have not been explored yet for a complex environment, i.e. saltmarsh ecosystem. Chapter 5 focused on this and evaluated the performance of SVM based on kernel selection and spectral separability indices. Chapter 5 also highlighted the application of small training size for wetland classification. Literature often suggests that the size of the training set required is a function of spectral wavebands used and generally a sample comprising at least 30 times the number of spectral bands used is required in the analysis process (Mather, 2004; Mathur and Foody 2008). That is why conventional classification scheme

(like MLC) require a large training set spread all over the study area. Based on this requirement, Worldview-2 data would require at least 240 pixels for each class to run a conventional MLC algorithm. Whereas, only 15% (16 pixels for each class) of the required training sample (240 pixels) provided up to 79% accuracy. This chapter also explored that optimal kernel choice is also highly variable among remote-sensing applications. This results also concur with other coastal vegetation studies where high-resolution data has been used. For example, Collin and Planes (2011) achieved more than 0.90 kappa coefficient with SVM from Worldview 2 data. Although the accuracy found in this chapter was lower than the one by Collin and Planes (2011), however, this lower accuracy was mainly due to our exclusive focus on individual species where spectral signatures overlapped. When Collins and Planes (2011) showed 28 pairwise comparisons in their land cover classification, only 1 showed a separability index inferior to 1.99. In our dataset, out of the 28 pairwise comparisons possible, 13 pairs displayed an inferior index to 1.99 (2.00 is a perfect discrimination) [See Appendix 1] and acted as an obstacle to improving accuracy after a certain degree.

Major limitations of this experiment were that parameters were randomly optimized instead of grid searching and cross-validation method. Feature selection is another important criterion of machine learning method application that was not considered in this experiment. In addition, the performance of SVM was not compared with any other ensemble classifiers to test the robustness of the SVM model. Finally, a sample size of this experiment was limited and fixed (16 pixels per class), therefore there was no scope to test the current results with a large sample size. To overcome these limitations, the author set another experiment that was discussed in chapter 6.

Ground truth data collection for species-level mapping is made challenging by limited access and hazardous conditions in some wetland ecosystems. Especially, saltmarsh classification is challenging in terms of a single (i.e water) and multi-modal distribution (seasonal variation of vegetation) features of saltmarsh habitat. Random Forest (RF) and Support Vector Machine (SVM) classifiers have been evaluated that are similar in their unique behaviour with multi-modal data distribution. However, there are some contradictory findings on the sensitivity of RF for sample size and labelling that are different from SVM. Therefore instead of area-based proportional sample allocation, an equal number of samples allocation strategies have been proposed for 8 classes within a saltmarshes community.

Feature or variable selection is a crucial issue in machine learning (RF and SVM) when dealing with advanced classification and regression problem (Hastie et al. 2001). To our knowledge, only a limited number of studies have been done, which explore advanced feature selection methods in multispectral data. Chapter 6 compared the performance of the relatively new advanced machine-learning RF and SVM classifiers on Worldview-2 imagery in a degraded saltmarsh wetland to determine the optimum sample size for classification and mapping purposes. This chapter contributes by answering the following research questions (1) How do machine learning algorithms of RF and SVM respond to the proportional reduction (100 % to 33%) of sample size? (2) Which machine learning approach provides reasonable accuracy with a limited number of samples? (3) Do the selected features of Worldview-2 change classification accuracy significantly compared to

using all bands of Worldview-2? In addition, this chapter highlights the parameter optimization process that was identified in chapter 5 as a major limitation.

The results of RF in chapter 6 showed a significant decrease in overall accuracy when the sample size was reduced from 100% level to 33%, the corresponding accuracy dropped from 79% to 72 %. Conversely, there were no significant changes in the accuracy of SVM, when samples size equally dropped from 100% to 33%. However, there was no significant difference among all the possible combinations of three experiments (100 %, 66% and 33% of the original samples) which originated from two classifiers based on an equal sample distribution scheme. The importance of each spectral bands was evaluated through Classification and Regression Training (CARET) and Random Forest (RF) packages in R scripting. Learning Vector Quantization (LVQ) and Recursive Feature Elimination (RFE), both feature selection methods selected six important features. No significant differences in classification accuracy were observed when these six important features were selected instead of all available features. This accuracy is promising and is comparable to other previous studies where multispectral and airborne hyperspectral data have been used. However, in terms of data acquisition cost, reference data collection and processing costs, this study showed a promising approach toward broad management application for the saltmarshes community that is ecologically endangered, and demand a special attention.

Besides mapping, there is a crucial need to quantify large-scale plant productivity, i.e. above ground biomass (AGB), in coastal marshes for a better understanding of marsh resilience against sea level rise (Schile et al. 2014; Swanson et al. 2014). But this AGB estimation is labour intensive and not feasible at large spatial extents. To complement this, remotely sensed data are utilized to map vegetation types and provide better estimates of plant production (Goetz and Dubayah 2011). Since the 1980s there have been some successful applications of remote sensing for saltmarsh biomass (Gross et al. 1987; Hardisky et al. 1983; Hardisky et al. 1984) using Normalized Difference Vegetation Index (NDVI). There are some major limitations of NDVI that have already been revealed by other studies (Gao et al. 2000; Tucker 1977). This studies recommended using narrow bands of hyperspectral data to overcome the limitations(Blackburn 1998; Thenkabail et al. 2000). However, there are some crucial limitations of hyperspectral data that we experienced for our study area and discussed in the introduction and first two chapters. To overcome these limitations, Worldview-2 with its higher spatial (1.84 m) and spectral (8 bands) resolution is seen as a tradeoff between the advantages of multispectral resolution satellite data and hyperspectral data (Mutanga et al. 2012; Rasel et al. 2016b).

Another research area that is still challenging is to model biomass against remote sensing spectral bands. Regression techniques are common to relate remotely sensed information with biophysical variables (i.e. biomass, leaf area index etc). However, they are also imitated to adequately capture this relationship and the spatiotemporal variability of the number of biophysical variables(Kaheil et al. 2008) . Moreover, multicollinearity is an important issue for the multiple regression model, especially when highly related variables (i.e bands of RS data) are selected as a predictor. An ensemble method, random forest (RF), has reduced the problem of the multicollinearity issue (Liaw and Wiener 2002a) and has been proved to reduce of bias and overfitting (Breiman 2001) of a model. Variable selection is a crucial issue in machine learning regression problem (Hastie et al. 2001). To

our knowledge, only a limited number of studies have been done, which explore these advanced feature selection methods in multispectral data (Pal 2006; Pal and Foody 2010). To address these problems, chapter 7 highlighted the performance of RF and SVM for saltmarsh biomass estimation using advance feature selection method embedded in RF and SVM.

Two machine learning approaches, Random Forest and Support Vector Machine regression were tested to estimate biomass of a common saltmarshes species, salt couch grass (Sporobolus virginicus). Reflectance and NDVI based vegetation indices derived from 8 bands of Worldview-2 multispectral data were used for four experiments to develop the biomass model. The main objectives of this study are (i) to recommend an affordable low-cost data source to predict biomass of a common saltmarshes species (ii) to suggest a variable selection method suitable for multispectral data (iii) to assess the performance of RF and SVM for the biomass prediction model. Cross-validation of parameter optimizations for SVM showed that the optimized parameter of ε-SVR failed to provide a reliable prediction. Hence, v-SVR was used for the SVM model. Among the different variable selection methods, Recursive Feature Elimination (RFE) selected a minimum number of variables (only 4) with an RMSE of 0.211 (kg/m2). Experiment-4 (only selected bands) provided the best results for both of the machine learning regression methods, RF (R2= 0.72, RMSE= 0.166 kg/m2) and SVR (R2= 0.66, RMSE = 0.200 kg/m2) to predict biomass. When a 10-fold cross validation of the RF model was compared with a 10-fold cross validation of SVR, a significant difference (p= <0.0001) was observed for RMSE. One to one comparisons of actual to predicted biomass showed that RF underestimates the high biomass values, whereas SVR overestimates the values; this suggests a need for further investigation and refinement.

The application of Random Forest and Support Vector Machine for predicting saltmarshes classification and biomass estimation at the species level is a new field of research and are not entirely investigated. However, the findings of this thesis enlighten the use of two machine learning algorithms for saltmarsh species classification and biomass estimation. This thesis focused on data comparison and data selection for a specific study area and then moved to the methodological approaches for classification and regression for saltmarshes monitoring.

In line with the application of Random Forest and variable selection methods, further studies could be undertaken along three lines: using airborne hyperspectral data collected from the same study area. Unnamed aerial vehicles (UAV) and any other hyperspectral image can be used. EnMAP, will be launched in 2019 and can be a good alternative for this. Another recommendation is the application of the same method and variable selection in the different study area with different species composition and considering the spatial autocorrelation of the samples for Random Forest application. Finally, different study area might follow the non-normal distribution pattern and will be useful to match the results of this research.

In this study, spaceborne hyperspectral data, Hyperion was compared with multispectral data. To make it compatible, the author selected only the very infrared part (VNIR) of the

EO-1 Hyperion that was similar with the wavelength of the multispectral Worldview-2 data. However, spatial resolution (30 m) of the EO-1 Hyperion was still an issue to detect saltmarsh species. Whereas, airborne hyperspectral data have very high spatial and spectral resolution which must be considered in saltmarsh mapping and biomass modelling. Although data acquisition is a challenging task due to several factors that have been discussed in chapter 2, 6 and 7, further research may be conducted if airborne hyperspectral data are available.

Saltmarsh species composition in the Hunter Wetland National park is limited in number and spatial size (small patches or fragmented). Only two major species, *Sporobolus virginicus* and *Phragmites australis* cohabit with other two tree species *Avicennia marina* and *Casuarina glauca*. Therefore the application of multispectral data Worldview-2, broadband data Landsat -8 OLI and hyperspectral data (either spaceborne or airborne) required to be investigated in a different study area with a complex vegetation composition. Although spaceborne hyperspectral data, EO-1 Hyperion has been decommissioned in 2017, another spaceborne hyperspectral data EnMAP will be launched in 2019 and can be considered in similar research.

In this study, Random Forest, a new ensemble classifier has been investigated for saltmarsh classification and biomass estimation. However, this classifier is sensitive to the spatial autocorrelation of the training classes (Millard and Richardson 2015) and imbalanced training classes (Dalponte et al. 2013) that have not been considered in this thesis. Further research is crucial to explore the sensitivity of this method with the spatial autocorrelation of the training samples and imbalanced allocation (different sample size for different classes) for random forest application in saltmarsh monitoring.

The results of this work, however, provide the necessary insight and motivation for wetland groups; ecologists, remote sensing group, machine learning users and environmentalists to shift toward the most affordable and easily accessible remote sensing sensors necessary for reliable aboveground biomass estimation for saltmarsh environment.

Appendix I:

Input File: WV_2_RST_Bil_Edited.tif

ROI Name: (Jeffries-Matusita, Transformed Divergence)

Mangrove [Red] 16 points:
- Casuarina [Green] 16 points: (1.84873406 1.98410535)
- Water [Blue] 16 points: (1.99985020 2.00000000)
- Phragmitis [Yellow] 16 points: (1.99873220 2.00000000)
- Sporobolus [Cyan] 16 points: (1.98635925 1.99997682)
- Wetland_Soil [Magenta] 16 points: (1.99982229 2.00000000)
- Dieback [Sea Green] 16 points: (1.84074828 1.99718994)
- Grass [Magenta] 16 points: (1.99420002 1.99999807)

Casuarina [Green] 16 points:
- Mangrove [Red] 16 points: (1.84873406 1.98410535)
- Water [Blue] 16 points: (1.97828754 2.00000000)
- Phragmitis [Yellow] 16 points: (1.99711086 2.00000000)
- Sporobolus [Cyan] 16 points: (1.98769264 1.99981308)
- Wetland_Soil [Magenta] 16 points: (1.99548826 1.99999999)
- Dieback [Sea Green] 16 points: (1.85521470 1.99967536)
- Grass [Magenta] 16 points: (1.99868982 2.00000000)

Water [Blue] 16 points:
- Mangrove [Red] 16 points: (1.99985020 2.00000000)
- Casuarina [Green] 16 points: (1.97828754 2.00000000)
- Phragmitis [Yellow] 16 points: (1.99984369 2.00000000)
- Sporobolus [Cyan] 16 points: (1.99809415 2.00000000)
- Wetland_Soil [Magenta] 16 points: (1.97902903 1.99999216)
- Dieback [Sea Green] 16 points: (1.98265840 2.00000000)
- Grass [Magenta] 16 points: (2.00000000 2.00000000)

Phragmitis [Yellow] 16 points:
- Mangrove [Red] 16 points: (1.99873220 2.00000000)
- Casuarina [Green] 16 points: (1.99711086 2.00000000)
- Water [Blue] 16 points: (1.99984369 2.00000000)
- Sporobolus [Cyan] 16 points: (1.94248532 1.99923084)
- Wetland_Soil [Magenta] 16 points: (1.99945312 2.00000000)
- Dieback [Sea Green] 16 points: (1.81660665 1.99913961)
- Grass [Magenta] 16 points: (1.99994145 2.00000000)

Sporobolus [Cyan] 16 points:
Mangrove [Red] 16 points: (1.98635925 1.99997682)
Casuarina [Green] 16 points: (1.98769264 1.99981308)
Water [Blue] 16 points: (1.99809415 2.00000000)
Phragmitis [Yellow] 16 points: (1.94248532 1.99923084)
Wetland_Soil [Magenta] 16 points: (1.99870232 1.99999999)
Dieback [Sea Green] 16 points: (1.48373069 1.87270344)
Grass [Magenta] 16 points: (1.99653838 1.99998930)

Wetland_Soil [Magenta] 16 points:
Mangrove [Red] 16 points: (1.99982229 2.00000000)
Casuarina [Green] 16 points: (1.99548826 1.99999999)
Water [Blue] 16 points: (1.97902903 1.99999216)
Phragmitis [Yellow] 16 points: (1.99945312 2.00000000)
Sporobolus [Cyan] 16 points: (1.99870232 1.99999999)
Dieback [Sea Green] 16 points: (1.97060942 1.99999982)
Grass [Magenta] 16 points: (2.00000000 2.00000000)

Dieback [Sea Green] 16 points:
Mangrove [Red] 16 points: (1.84074828 1.99718994)
Casuarina [Green] 16 points: (1.85521470 1.99967536)
Water [Blue] 16 points: (1.98265840 2.00000000)
Phragmitis [Yellow] 16 points: (1.81660665 1.99913961)
Sporobolus [Cyan] 16 points: (1.48373069 1.87270344)
Wetland_Soil [Magenta] 16 points: (1.97060942 1.99999982)
Grass [Magenta] 16 points: (1.97437798 1.99927483)

Grass [Magenta] 16 points:
Mangrove [Red] 16 points: (1.99420002 1.99999807)
Casuarina [Green] 16 points: (1.99868982 2.00000000)
Water [Blue] 16 points: (2.00000000 2.00000000)
Phragmitis [Yellow] 16 points: (1.99994145 2.00000000)
Sporobolus [Cyan] 16 points: (1.99653838 1.99998930)
Wetland_Soil [Magenta] 16 points: (2.00000000 2.00000000)
Schoeloplectus [Sea Green] 16 points: (1.97437798 1.99927483)

Pair Separation (least to most);

Sporobolus [Cyan] 16 points and Dieback [Sea Green] 16 points - 1.48373069
Phragmitis [Yellow] 16 points and Dieback[Sea Green] 16 points - 1.81660665
Mangrove [Red] 16 points and Dieback[Sea Green] 16 points - 1.84074828
Mangrove [Red] 16 points and Casuarina [Green] 16 points - 1.84873406

Casuarina [Green] 16 points and Dieback [Sea Green] 16 points - 1.85521470
Phragmitis [Yellow] 16 points and Sporobolus [Cyan] 16 points - 1.94248532
Wetland_Soil [Magenta] 16 points and Dieback[Sea Green] 16 points - 1.97060942
Dieback [Sea Green] 16 points and Grass [Magenta] 16 points - 1.97437798
Casuarina [Green] 16 points and Water [Blue] 16 points - 1.97828754
Water [Blue] 16 points and Wetland_Soil [Magenta] 16 points - 1.97902903
Water [Blue] 16 points and Dieback [Sea Green] 16 points - 1.98265840
Mangrove [Red] 16 points and Sporobolus [Cyan] 16 points - 1.98635925
Casuarina [Green] 16 points and Sporobolus [Cyan] 16 points - 1.98769264
Mangrove [Red] 16 points and Grass [Magenta] 16 points - 1.99420002
Casuarina [Green] 16 points and Wetland_Soil [Magenta] 16 points - 1.99548826
Sporobolus [Cyan] 16 points and Grass [Magenta] 16 points - 1.99653838
Casuarina [Green] 16 points and Phragmitis [Yellow] 16 points - 1.99711086
Water [Blue] 16 points and Sporobolus [Cyan] 16 points - 1.99809415
Casuarina [Green] 16 points and Grass [Magenta] 16 points - 1.99868982
Sporobolus [Cyan] 16 points and Wetland_Soil [Magenta] 16 points - 1.99870232
Mangrove [Red] 16 points and Phragmitis [Yellow] 16 points - 1.99873220
Phragmitis [Yellow] 16 points and Wetland_Soil [Magenta] 16 points - 1.99945312
Mangrove [Red] 16 points and Wetland_Soil [Magenta] 16 points - 1.99982229
Water [Blue] 16 points and Phragmitis [Yellow] 16 points - 1.99984369
Mangrove [Red] 16 points and Water [Blue] 16 points - 1.99985020
Phragmitis [Yellow] 16 points and Grass [Magenta] 16 points - 1.99994145
Wetland_Soil [Magenta] 16 points and Grass [Magenta] 16 points - 2.00000000
Water [Blue] 16 points and Grass [Magenta] 16 points - 2.000000

Bibliography

Adam, E., Mutanga, O., Odindi, J., & Abdel-Rahman, E.M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing, 35*, 3440-3458

Adam, E., Mutanga, O., & Rugege, D. (2010). Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands Ecology and Management, 18*, 281-296

Adam, P. (1993). *Saltmarsh ecology*. Cambridge University Press

Adelabu, S., Mutanga, O., & Adam, E. (2014). Evaluating the impact of red-edge band from Rapideye image for classifying insect defoliation levels. *ISPRS Journal of Photogrammetry and Remote Sensing, 95*, 34-41

Almeida, T., & Filho, D.S. (2004). Principal component analysis applied to feature-oriented band ratios of hyperspectral data: a tool for vegetation studies. *International Journal of Remote Sensing, 25*, 5005-5023

Amani, M., Salehi, B., Mahdavi, S., Granger, J.E., Brisco, B., & Hanson, A. (2017). Wetland Classification Using Multi-Source and Multi-Temporal Optical Remote Sensing Data in Newfoundland and Labrador, Canada. *Canadian Journal of Remote Sensing, 43*, 360-373

Archibald, R., & Fann, G. (2007). Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geoscience and Remote Sensing Letters, 4*, 674-677

Arora, M., & Foody, G. (1997). Log-linear modelling for the evaluation of the variables affecting the accuracy of probabilistic, fuzzy and neural network classifications. *International Journal of Remote Sensing, 18*, 785-798

Atkinson, P., Sargent, I., Foody, G., & Williams, J. (2007). Exploring the geostatistical method for estimating the signal-to-noise ratio of images. *Photogrammetric Engineering & Remote Sensing, 73*, 841-850

Axelsson, C., Skidmore, A.K., Schlerf, M., Fauzi, A., & Verhoef, W. (2013). Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. *International Journal of Remote Sensing, 34*, 1724-1743

Bazi, Y., & Melgani, F. (2006). Toward an optimal SVM classification system for hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 44*, 3374-3385

Beck, R. (2003). EO-1 user guide, version 2.3. *Satellite Systems Branch, USGS Earth Resources Observation Systems Data Center (EDC)*

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24-31

Belluco, E., Camuffo, M., Ferrari, S., Modenese, L., Silvestri, S., Marani, A., & Marani, M. (2006). Mapping salt-marsh vegetation by multispectral and hyperspectral remote sensing. *Remote sensing of environment, 105*, 54-67

Ben-Hur, A., & Weston, J. (2010). A user's guide to support vector machines. *Data mining techniques for the life sciences*, 223-239

Blackburn, G.A. (1998). Quantifying chlorophylls and caroteniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote sensing of environment, 66*, 273-285

Breiman, L. (2001). Random forests. *Machine learning, 45*, 5-32

Brown, M., Lewis, H.G., & Gunn, S.R. (2000). Linear spectral mixture models and support vector machines for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing, 38*, 2346-2360

Bruzzone, L., Chi, M., & Marconcini, M. (2006). A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 44*, 3363-3373

Bruzzone, L., & Melgani, F. (2005). Robust multiple estimator systems for the analysis of biophysical parameters from remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing, 43*, 159-174

Burges, C.J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery, 2*, 121-167

Byrd, K.B., O'Connell, J.L., Di Tommaso, S., & Kelly, M. (2014). Evaluation of sensor types and environmental controls on mapping biomass of coastal marsh emergent vegetation. *Remote sensing of environment, 149*, 166-180

Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Martín-Guerrero, J.D., Soria-Olivas, E., Alonso-Chordá, L., & Moreno, J. (2004). Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 1530-1542

Carle, M.V., Wang, L., & Sasser, C.E. (2014). Mapping freshwater marsh species distributions using WorldView-2 high-resolution multispectral satellite imagery. *International Journal of Remote Sensing, 35*, 4698-4716

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., & Woodcock, C.E. (1999). A neural network method for mixture estimation for vegetation mapping. *Remote sensing of environment, 70*, 138-152

Castillo, C., Chollett, I., & Klein, E. (2008). Enhanced duckweed detection using bootstrapped SVM classification on medium resolution RGB MODIS imagery. *International Journal of Remote Sensing, 29*, 5595-5604

Cavallaro, G., Dalla Mura, M., Benediktsson, J.A., & Plaza, A. (2016). Remote sensing image classification using attribute filters defined over the tree of shapes. *IEEE Transactions on Geoscience and Remote Sensing, 54*, 3899-3911

Chan, J.C.-W., & Paelinckx, D. (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*, 2999-3011

Chauhan, H.J., Mohan, B.K., & CSRE, I. (2011). Approaches to Optimum Dimensions Selection of Remotely Sensed EO-1 Hyperion Data for Crop Classification. *IJCSI*, 12

Chen, C.H., & Ho, P.-G.P. (2008). Statistical pattern recognition in remote sensing. *Pattern Recognition, 41*, 2731-2741

Chen, D., Stow, D., & Gong, P. (2004). Examining the effect of spatial resolution and texture window size on classification accuracy: an urban environment case. *International Journal of Remote Sensing, 25*, 2177-2192

Chen, G., & Hay, G.J. (2011). A support vector regression approach to estimate forest biophysical parameters at the object level using airborne lidar transects and QuickBird data. *Photogrammetric Engineering & Remote Sensing, 77*, 733-741

Chen, J., Gu, S., Shen, M., Tang, Y., & Matsushita, B. (2009). Estimating aboveground biomass of grassland having a high canopy cover: an exploratory analysis of in situ hyperspectral data. *International Journal of Remote Sensing, 30*, 6497-6517

Cherkassky, V., Shao, X., Mulier, F.M., & Vapnik, V.N. (1999). Model complexity control for regression using VC generalization bounds. *IEEE transactions on Neural Networks, 10*, 1075-1089

Chi, M., Feng, R., & Bruzzone, L. (2008). Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Advances in space research, 41*, 1793-1799

Chmura, G.L. (2009). Tidal salt marshes. *The management of natural coastal carbon sinks*, 5

Chmura, G.L., Anisfeld, S., Cahoon, D.a., & J, L. (2003). Global carbon sequestration in tidal, saline wetland soils. . *Global Biogeochemical cycles, 17*, 1-12

Cho, M.A., Skidmore, A., Corsi, F., Van Wieren, S.E., & Sobhan, I. (2007). Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *International Journal of Applied Earth Observation and Geoinformation, 9*, 414-424

Clevers, J., Van der Heijden, G., Verzakov, S., & Schaepman, M. (2007). Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogrammetric Engineering & Remote Sensing, 73*, 1141-1148

Colditz, R.R. (2015). An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. *Remote Sensing, 7*, 9655-9681

Collin, A., & Planes, S. (2011). What is the value added of 4 bands within the submetric remote sensing of tropical coastscape? Quickbird-2 vs WorldView-2. In, *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International* (pp. 2165-2168): IEEE

Congalton, R., & Green, K. (1999). Assessing the accuracy of remotely sensed data: principles and applications. *Lewis Pub—lishers, Boca Raton, Fla*

Congalton, R.G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*, 273-297

Curran, P.J., & Atkinson, P.M. (1999). Issues of scale and optimal pixel size. *Spatial statistics for remote sensing* (pp. 115-133): Springer

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., & Lawler, J.J. (2007). Random forests for classification in ecology. *Ecology, 88*, 2783-2792

Dalponte, M., Bruzzone, L., & Gianelle, D. (2008). Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas. *IEEE Transactions on Geoscience and Remote Sensing, 46*, 1416-1427

Dalponte, M., Orka, H.O., Gobakken, T., Gianelle, D., & Næsset, E. (2013). Tree species classification in boreal forests with hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing, 51*, 2632-2645

Daly, T. (2013). Prime fact: Coastal Saltmarsh. In. NSW , Australia

Datt, B., McVicar, T.R., Van Niel, T.G., Jupp, D.L., & Pearlman, J.S. (2003). Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *Geoscience and Remote Sensing, IEEE Transactions on, 41*, 1246-1259

Davranche, A., Lefebvre, G., & Poulin, B. (2010). Wetland monitoring using classification trees and SPOT-5 seasonal time series. *Remote sensing of environment, 114*, 552-562

Demarchi, L., Canters, F., Cariou, C., Licciardi, G., & Chan, J.C.-W. (2014). Assessing the performance of two unsupervised dimensionality reduction techniques on hyperspectral APEX data for high resolution urban land-cover mapping. *ISPRS Journal of Photogrammetry and Remote Sensing, 87*, 166-179

Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics, 7*

Díaz-Uriarte, R., & De Andres, S.A. (2006a). Gene selection and classification of microarray data using random forest. *BMC bioinformatics, 7*, 1
Díaz-Uriarte, R., & De Andres, S.A. (2006b). Gene selection and classification of microarray data using random forest. *BMC bioinformatics, 7*, 3
DoF (2010). Hamro Ban, Community Forest Division, Department of forests, Kathmundu, Nepal.
Du, P., Samat, A., Waske, B., Liu, S., & Li, Z. (2015). Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 38-53
Dube, T., & Mutanga, O. (2015). Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing, 101*, 36-46
Duda, R.O., Hart, P.E., & Stork, D.G. (2012). *Pattern classification*. John Wiley & Sons
Durbha, S.S., King, R.L., & Younan, N.H. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote sensing of environment, 107*, 348-361
Duro, D.C., Franklin, S.E., & Dubé, M.G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote sensing of environment, 118*, 259-272
Eckert, S. (2012). Improved forest biomass and carbon estimations using texture measures from WorldView-2 satellite data. *Remote Sensing, 4*, 810-829
Elsaid, M., Aboelkhair, H., Dardier, A., & Hermas, E. (2014). Investigation of a relation between radiogenic heat production and kinetic surface temperature from multispectral ASTER-TIR data: a case study on Elmissikat-Eleridiya granites, Central Eastern Desert, Egypt. *Arabian Journal of Geosciences, 7*, 4615-4628
Englhart, S., Keuck, V., & Siegert, F. (2012). Modeling aboveground biomass in tropical forests using multi-frequency SAR data—A comparison of methods. *IEEE Journal of selected topics in applied earth observations and remote sensing, 5*, 298-306
Fang, C., Tao, Z., Gao, D., & Wu, H. (2016). Wetland mapping and wetland temporal dynamic analysis in the Nanjishan wetland using Gaofen One data. *Annals of GIS, 22*, 259-271
Foody, G.M. (2008). RVM-based multi-class classification of remotely sensed data. *International Journal of Remote Sensing, 29*, 1817-1823
Foody, G.M., & Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on, 42*, 1335-1343
Foody, G.M., & Mathur, A. (2006). The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote sensing of environment, 103*, 179-189
Foody, G.M., Mathur, A., Sanchez-Hernandez, C., & Boyd, D.S. (2006). Training set size requirements for the classification of a specific class. *Remote sensing of environment, 104*, 1-14
Franklin, S.E., & Ahmed, O.S. (2017). Object-based Wetland Characterization Using Radarsat-2 Quad-Polarimetric SAR Data, Landsat-8 OLI Imagery, and Airborne Lidar-Derived Geomorphometric Variables. *Photogrammetric Engineering & Remote Sensing, 83*, 27-36

Frazier, R.J., Coops, N.C., Wulder, M.A., & Kennedy, R. (2014). Characterization of aboveground biomass in an unmanaged boreal forest using Landsat temporal segmentation metrics. *ISPRS Journal of Photogrammetry and Remote Sensing, 92*, 137-146

Fu, B., Wang, Y., Campbell, A., Li, Y., Zhang, B., Yin, S., Xing, Z., & Jin, X. (2017). Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecological Indicators, 73*, 105-117

Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Academic press

FUKUNAGA, R. (1990). Statistical pattern recognition

Gadallah, F., & Csillag, F. (2002). Vegetation quantification in a sub-arctic salt marsh using reflectance data. In, *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International* (pp. 3293-3295): IEEE

Gao, X., Huete, A.R., Ni, W., & Miura, T. (2000). Optical–biophysical relationships of vegetation spectra without background contamination. *Remote sensing of environment, 74*, 609-620

Gedan, K.B., Silliman, B.R., & Bertness, M.D. (2009). Centuries of human-driven change in salt marsh ecosystems. *Annual review of marine science, 1*, 117-141

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*, 2225-2236

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2015). VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal, 7*, 19-33

Ghoggali, N., Melgani, F., & Bazi, Y. (2009). A multiobjective genetic SVM approach for classification problems with limited training samples. *IEEE Transactions on Geoscience and Remote Sensing, 47*, 1707-1718

Ghosh, A., & Joshi, P.K. (2014). A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery. *International Journal of Applied Earth Observation and Geoinformation, 26*, 298-311

Gislason, P.O., Benediktsson, J.A., & Sveinsson, J.R. (2006). Random forests for land cover classification. *Pattern Recognition Letters, 27*, 294-300

Gleason, C.J., & Im, J. (2012). Forest biomass estimation from airborne LiDAR data using machine learning approaches. *Remote sensing of environment, 125*, 80-91

Globe, D. (2010). The benefit of the eight spectral bands of Worldview-2. In: Digital Globe

Goetz, S., & Dubayah, R. (2011). Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change. *Carbon Management, 2*, 231-244

Green, A.A., Berman, M., Switzer, P., & Craig, M.D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *Geoscience and Remote Sensing, IEEE Transactions on, 26*, 65-74

Gross, M., Hardisky, M., Klemas, V., & Wolf, P. (1987). Quantification of biomass of the marsh grass Spartina alterniflora Loisel using Landsat Thematic Mapper imagery. *Photogrammetric engineering and remote sensing, 53*, 1577-1583

Gross, M., Klemas, V., & Levasseur, J. (1986). Remote sensing of Spartina anglica biomass in five French salt marshes. *International Journal of Remote Sensing, 7*, 657-664

Gross, M.F., Hardisky, M.A., Wolf, P.L., & Klemas, V. (1991). Relationship between aboveground and belowground biomass ofSpartina alterniflora (smooth cordgrass). *Estuaries, 14*, 180-191

Gualtieri, J.A., & Cromp, R.F. (1998). Support vector machines for hyperspectral remote sensing classification

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*, 1157-1182

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

Haas, J., & Ban, Y. (2014). Urban growth and environmental impacts in Jing-Jin-Ji, the Yangtze, River Delta and the Pearl River Delta. *International Journal of Applied Earth Observation and Geoinformation, 30*, 42-55

Ham, J., Chen, Y., Crawford, M.M., & Ghosh, J. (2005). Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing, 43*, 492-501

Han, T., Goodenough, D., Dyk, A., & Love, J. (2002). Detection and correction of abnormal pixels in Hyperion images. In, *Geoscience and Remote Sensing Symposium, 2002. IGARSS'02. 2002 IEEE International* (pp. 1327-1330): IEEE

Hanselmann, M., Köthe, U., Kirchner, M., Renard, B.Y., Amstalden, E.R., Glunde, K., Heeren, R.M., & Hamprecht, F.A. (2009). Towards Digital Staining using Imaging Mass Spectrometry and Random Forests-Technical Report

Hardisky, M., Smart, R., & Klemas, V. (1983). Seasonal spectral characteristics and aboveground biomass of the tidal marsh plant, Spartina alterniflora. *Photogrammetric engineering and remote sensing, 49*, 85-92

Hardisky, M.A., Daiber, F.C., Roman, C.T., & Klemas, V. (1984). Remote sensing of biomass and annual net aerial primary productivity of a salt marsh. *Remote sensing of environment, 16*, 91-106

Harvey, K., & Hill, G. (2001). Vegetation mapping of a tropical freshwater swamp in the Northern Territory, Australia: a comparison of aerial photography, Landsat TM and SPOT satellite imagery. *International Journal of Remote Sensing, 22*, 2911-2925

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning. 2001. *NY Springer*

He, C., Zhang, Q., Li, Y., Li, X., & Shi, P. (2005). Zoning grassland protection area using remote sensing and cellular automata modeling—a case study in Xilingol steppe grassland in northern China. *Journal of Arid Environments, 63*, 814-826

Herold, M., Roberts, D.A., Gardner, M.E., & Dennison, P.E. (2004). Spectrometry for urban area remote sensing—Development and analysis of a spectral library from 350 to 2400 nm. *Remote sensing of environment, 91*, 304-319

Horler, D., DOCKRAY, M., & Barber, J. (1983). The red edge of plant leaf reflectance. *International Journal of Remote Sensing, 4*, 273-288

Hornik, K., Meyer, D., & Karatzoglou, A. (2006). Support vector machines in R. *Journal of Statistical Software, 15*, 1-28

Huang, C., Davis, L., & Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing, 23*, 725-749

Huang, X., & Zhang, L. (2013). An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing, 51*, 257-272

Hughes, G.P. (1968). On the mean accuracy of statistical pattern recognizers. *Information Theory, IEEE Transactions on, 14*, 55-63
Hunter, E., & Power, C. (2002). An assessment of two classification methods for mapping Thames Estuary intertidal habitats using CASI data. *International Journal of Remote Sensing, 23*, 2989-3008
Ismail, R., Mutanga, O., & Kumar, L. (2010). Modeling the potential distribution of pine forests susceptible to sirex noctilio infestations in Mpumalanga, South Africa. *Transactions in GIS, 14*, 709-726
Jackson, R.D., & Huete, A.R. (1991). Interpreting vegetation indices. *Preventive veterinary medicine, 11*, 185-200
Jensen, J.R., Olson, G., Schill, S.R., Porter, D.E., & Morris, J. (2002). Remote Sensing of Biomass, Leaf-Area-Index, and Chlorophyll a and b Content in the ACE Basin National Estuarine Research Reserve Using Sub-meter Digital Camera Imagery. *Geocarto international, 17*, 27-36
Jia, X., & Richards, J.A. (1999). Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Transactions on Geoscience and Remote Sensing, 37*, 538-542
Jiang, X., Tang, L., Wang, C., & Wang, C. (2004). Spectral characteristics and feature selection of hyperspectral remote sensing data. *International Journal of Remote Sensing, 25*, 51-59
Jin, H., Stehman, S.V., & Mountrakis, G. (2014). Assessing the impact of training sample selection on accuracy of an urban classification: a case study in Denver, Colorado. *International Journal of Remote Sensing, 35*, 2067-2081
Jin, S., Li, D., & Gong, J. (2005). A comparison of SVMs with MLC algorithms on texture features. In (pp. 60442B-60442B-60446)
Johnston, R., & Barson, M. (1993). Remote sensing of Australian wetlands: An evaluation of Landsat TM data for inventory and classification. *Marine and Freshwater Research, 44*, 235-252
Jupp, D.L.B., & Datt, B. (2004). *Evaluation of the EO-1 Hyperion Hyperspectral Instrument & its Applications at Australian Validation Sites 2001-2003*. CSIRO Earth Observation Centre
Kaheil, Y.H., Rosero, E., Gill, M.K., McKee, M., & Bastidas, L.A. (2008). Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 46*, 2692-2707
Karlson, M., Ostwald, M., Reese, H., Sanou, J., Tankoano, B., & Mattsson, E. (2015). Mapping tree canopy cover and aboveground biomass in Sudano-Sahelian woodlands using Landsat 8 and random forest. *Remote Sensing, 7*, 10017-10041
Kaufman, Y., Tanré, D., Holben, B., Markham, B., & Gitelson, A.A. (1992). Atmospheric Effects on the NDVI--Strategies for its Removal
Kavzoglu, T., & Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation, 11*, 352-359
Kavzoglu, T., & Reis, S. (2008). Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels. *GIScience & Remote Sensing, 45*, 330-342
Keramitsoglou, I., Sarimveis, H., Kiranoudis, C.T., Kontoes, C., Sifakis, N., & Fitoka, E. (2006). The performance of pixel window algorithms in the classification of habitats

using VHSR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 60*, 225-238
Kloiber, S.M., Macleod, R.D., Smith, A.J., Knight, J.F., & Huberty, B.J. (2015). A semi-automated, multi-source data fusion update of a wetland inventory for east-central Minnesota, USA. *Wetlands, 35*, 335-348
Knorn, J., Rabe, A., Radeloff, V.C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote sensing of environment, 113*, 957-964
Kohavi, R., & John, G.H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*, 273-324
Kruse, F., Boardman, J.W., & Huntington, J.F. (2003). Comparison of airborne hyperspectral data and EO-1 Hyperion for mineral mapping. *Geoscience and Remote Sensing, IEEE Transactions on, 41*, 1388-1400
Kuhn, M. (2008). Caret package. *Journal of Statistical Software, 28*, 1-26
Kumar, L., & Sinha, P. (2014). Mapping salt-marsh land-cover vegetation using high-spatial and hyperspectral satellite data to assist wetland inventory. *GIScience & Remote Sensing, 51*, 483-497
Kursa, M.B., & Rudnicki, W.R. (2010). Feature selection with the Boruta package. In: Journal
Kwiatkowska, E.J., & Fargion, G.S. (2003). Application of machine-learning techniques toward the creation of a consistent and calibrated global chlorophyll concentration baseline dataset using remotely sensed ocean color data. *IEEE Transactions on Geoscience and Remote Sensing, 41*, 2844-2860
Lardeux, C., Frison, P.-L., Tison, C., Souyris, J.-C., Stoll, B., Fruneau, B., & Rudant, J.-P. (2009). Support vector machine for multifrequency SAR polarimetric data classification. *IEEE Transactions on Geoscience and Remote Sensing, 47*, 4143-4152
Lavergne, P., & Patilea, V. (2008). Breaking the curse of dimensionality in non parametric testing. *Journal of Econometrics, 143*, 103-122
Lawrence, R.L., & Ripple, W.J. (1998). Comparisons among vegetation indices and bandwise regression in a highly disturbed, heterogeneous landscape: Mount St. Helens, Washington. *Remote sensing of environment, 64*, 91-102
Lawrence, R.L., Wood, S.D., & Sheley, R.L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote sensing of environment, 100*, 356-362
Lee, K., & Lunetta, R.S. (1995). Wetland detection methods. *JG Lyon, and J. McCarthy. Eds, Wetland and environmental applications in GIS. Lewis, Boca Raton, Florida*, 249-284
Leutner, B.F., Reineking, B., Müller, J., Bachmann, M., Beierkuhnlein, C., Dech, S., & Wegmann, M. (2012). Modelling forest α-diversity and floristic composition—On the added value of LiDAR plus hyperspectral remote sensing. *Remote Sensing, 4*, 2818-2845
Li, H., Gu, H., Han, Y., & Yang, J. (2010). Object-oriented classification of high-resolution remote sensing imagery based on an improved colour structure code and a support vector machine. *International Journal of Remote Sensing, 31*, 1453-1470
Li, L., Ustin, S., & Lay, M. (2005). Application of multiple endmember spectral mixture analysis (MESMA) to AVIRIS imagery for coastal salt marsh mapping: a case study in China Camp, CA, USA. *International Journal of Remote Sensing, 26*, 5193-5207

Li, X., Chen, W., Cheng, X., Liao, Y., & Chen, G. (2017). Comparison and integration of feature reduction methods for land cover classification with RapidEye imagery. *Multimedia Tools and Applications*

Li, Z., Xu, D., & Guo, X. (2014). Remote sensing of ecosystem health: opportunities, challenges, and future perspectives. *Sensors, 14*, 21117-21139

Liaw, A. (2002). Classification and Regression by randomForest. *R News, 2*, 18-22

Liaw, A., & Wiener, M. (2002a). Classification and Regression by randomForest. *R News, 2*

Liaw, A., & Wiener, M. (2002b). Classification and Regression by randomForest

Liu, G., Zhou, D., Xu, H., & Mei, C. (2010). Model optimization of SVM for a fermentation soft sensor. *Expert Systems with Applications, 37*, 2708-2713

Löw, F., Michel, U., Dech, S., & Conrad, C. (2013). Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing, 85*, 102-119

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing, 28*, 823-870

Lu, F., & Petkova, E. (2014). A comparative study of variable selection methods in the context of developing psychiatric screening instruments. *Statistics in medicine, 33*, 401-421

Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., Ma, X., & Chen, D. (2017). Evaluation of Feature Selection Methods for Object-Based Land Cover Mapping of Unmanned Aerial Vehicle Imagery Using Random Forest and Support Vector Machine Classifiers. *ISPRS International Journal of Geo-Information, 6*, 51

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., & Motagh, M. (2017). Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 130*, 13-31

Mansour, K., Mutanga, O., Everson, T., & Adam, E. (2012). Discriminating indicator grass species for rangeland degradation assessment using hyperspectral data resampled to AISA Eagle resolution. *ISPRS Journal of Photogrammetry and Remote Sensing, 70*, 56-65

Mantero, P., Moser, G., & Serpico, S.B. (2005). Partially supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Transactions on Geoscience and Remote Sensing, 43*, 559-570

Marabel, M., & Alvarez-Taboada, F. (2013). Spectroscopic determination of aboveground biomass in grasslands using spectral transformations, support vector machine and partial least squares regression. *Sensors, 13*, 10027-10051

Markham, B., & Townshend, J. (1981). Land cover classification accuracy as a function of sensor spatial resolution

Marshall, V., Lewis, M., & Ostendorf, B. (2012). Do additional bands (coastal, NIR-2, red-edge and yellow) in WorldView-2 multispectral imagery improve discrimination of an Invasive Tussock, Buffel Grass (Cenchrus Ciliaris). *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 39*, B8

Mas, J.F., & Flores, J.J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing, 29*, 617-663

Mascaro, J., Asner, G.P., Knapp, D.E., Kennedy-Bowdoin, T., Martin, R.E., Anderson, C., Higgins, M., & Chadwick, K.D. (2014). A tale of two "forests": Random Forest machine learning aids tropical forest carbon mapping. *PloS one, 9*, e85993

Mathur, A., & Foody, G. (2008a). Multiclass and binary SVM classification: Implications for training and classification users. *Geoscience and Remote Sensing Letters, IEEE, 5*, 241-245

Mathur, A., & Foody, G.M. (2008b). Multiclass and binary SVM classification: Implications for training and classification users. *IEEE Geoscience and Remote Sensing Letters, 5*, 241-245

Mazzoni, D., Horváth, Á., Garay, M.J., Tang, B., & Davies, R. (2005). A MISR cloud-type classifier using reduced Support Vector Machines

Mehmood, T., Liland, K.H., Snipen, L., & Sæbø, S. (2012). A review of variable selection methods in partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 118*, 62-69

Mellor, A., Boukir, S., Haywood, A., & Jones, S. (2015). Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 155-168

Middleton, E.M. (1991). Solar zenith angle effects on vegetation indices in tallgrass prairie. *Remote sensing of environment, 38*, 45-62

Millard, K., & Richardson, M. (2013). Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier. *Canadian Journal of Remote Sensing, 39*, 290-307

Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in random forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing, 7*, 8489-8515

Montes, J., Technow, F., Dhillon, B., Mauch, F., & Melchinger, A. (2011). High-throughput non-destructive biomass determination during early plant development in maize under field conditions. *Field Crops Research, 121*, 268-273

Morris, J.T., Sundareshwar, P., Nietch, C.T., Kjerfve, B., & Cahoon, D.R. (2002). Responses of coastal wetlands to rising sea level. *Ecology, 83*, 2869-2877

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*, 247-259

Mutanga, O., Adam, E., & Cho, M.A. (2012). High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation, 18*, 399-406

Mutanga, O., & Skidmore, A. (2004a). Integrating imaging spectroscopy and neural networks to map grass quality in the Kruger National Park, South Africa. *Remote sensing of environment, 90*, 104-115

Mutanga, O., & Skidmore, A.K. (2004b). Hyperspectral band depth analysis for a better estimation of grass biomass (Cenchrus ciliaris) measured under controlled laboratory conditions. *International Journal of Applied Earth Observation and Geoinformation, 5*, 87-96

Mutanga, O., Skidmore, A.K., & Prins, H. (2004). Predicting in situ pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote sensing of environment, 89*, 393-408

Neumann, M., Saatchi, S.S., Ulander, L.M., & Fransson, J.E. (2012). Assessing performance of L-and P-band polarimetric interferometric SAR data in estimating boreal

forest above-ground biomass. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 714-726
Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proc. of the 4th GEOBIA*, 7-9
Olden, J.D., Lawler, J.J., & Poff, N.L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology, 83*, 171-193
Omar, H. (2010). Commercial timber tree species identification using multispectral Worldview2 data. *Digital Globe® 8Bands Research Challenge*, 2-13
Oommen, T., Misra, D., Twarakavi, N.K., Prakash, A., Sahoo, B., & Bandopadhyay, S. (2008). An objective analysis of support vector machine based classification for remote sensing. *Mathematical geosciences, 40*, 409-424
Otukei, J.R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation, 12*, S27-S31
Ouyang, Z.-T., Zhang, M.-Q., Xie, X., Shen, Q., Guo, H.-Q., & Zhao, B. (2011). A comparison of pixel-based and object-oriented approaches to VHR imagery for mapping saltmarsh plants. *Ecological informatics, 6*, 136-146
Øystein, R., & Øivind, D.T. (2012). Evaluation of FLAASH ATMOSPHERIC CORRECTION. In (p. 24): Norwegian Comuting Centre
Ozesmi, S.L., & Bauer, M.E. (2002). Satellite remote sensing of wetlands. *Wetlands Ecology and Management, 10*, 381-402
Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*, 217-222
Pal, M. (2006). Support vector machine-based feature selection for land cover classification: a case study with DAIS hyperspectral data. *International Journal of Remote Sensing, 27*, 2877-2894
Pal, M. (2008). Ensemble of support vector machines for land cover classification. *International Journal of Remote Sensing, 29*, 3043-3049
Pal, M., & Foody, G.M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing, 48*, 2297-2307
Pal, M., & Mather, P. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing, 26*, 1007-1011
Pal, M., & Mather, P.M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment, 86*, 554-565
Pearlman, J.S., Barry, P.S., Segal, C.C., Shepanski, J., Beiso, D., & Carman, S.L. (2003). Hyperion, a space-based imaging spectrometer. *Geoscience and Remote Sensing, IEEE Transactions on, 41*, 1160-1173
Peerbhay, K., Mutanga, O., Lottering, R., & Ismail, R. (2016). Mapping Solanum mauritianum plant invasions using WorldView-2 imagery and unsupervised random forests. *Remote sensing of environment, 182*, 39-48
Pervez, W., & Khan, S. (2015). Hyperspectral Hyperion Imagery Analysis and its Application Using Spectral Analysis. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 1*, 169-175
Peters, A., Hothorn, T., & Lausen, B. (2002). ipred: Improved predictors. *R News, 2*, 33-36

Petropoulos, G.P., Kalaitzidis, C., & Vadrevu, K.P. (2012). Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Computers & Geosciences, 41*, 99-107

Pouteau, R., Meyer, J.-Y., Taputuarai, R., & Stoll, B. (2012). Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological informatics, 9*, 37-46

Powell, S.L., Cohen, W.B., Healey, S.P., Kennedy, R.E., Moisen, G.G., Pierce, K.B., & Ohmann, J.L. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote sensing of environment, 114*, 1053-1068

Proisy, C., Couteron, P., & Fromard, F. (2007). Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. *Remote sensing of environment, 109*, 379-392

Qi, J., Moran, M., Cabot, F., & Dedieu, G. (1995). Normalization of sun/view angle effects using spectral albedo-based vegetation indices. *Remote sensing of environment, 52*, 207-217

Qu, L., Han, W., Lin, H., Zhu, Y., & Zhang, L. (2014). Estimating vegetation fraction using hyperspectral pixel unmixing method: a case study of a karst area in China. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of, 7*, 4559-4565

Räsänen, A., Rusanen, A., Kuitunen, M., & Lensu, A. (2013). What makes segmentation good? A case study in boreal forest habitat mapping. *International Journal of Remote Sensing, 34*, 8603-8627

Rasel, S., Groen, T., Hussin, Y., & Diti, I. (2017). Proxies for soil organic carbon derived from remote sensing. *International Journal of Applied Earth Observation and Geoinformation, 59*, 157-166

Rasel, S.M., Chang, H.-C., Diti, I.J., Ralph, T., & Saintilan, N. (2016a). Comparative analysis of Worldview-2 and Landsat 8 for coastal saltmarsh mapping accuracy assessment. In, *Sensing for Agriculture and Food Quality and Safety VIII*

Rasel, S.M.M., Chang, H.-C., Diti, I.J., Ralph, T., & Saintilan, N. (2016b). COMPARISON OF VERY NEAR INFRARED (VNIR) WAVELENGTH FROM EO-1 HYPERION AND WORLDVIEW 2 IMAGES FOR SALTMARSH CLASSIFICATION. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 3*

Rendong, L., & Jiyuan, L. (2004). Estimating wetland vegetation biomass in the Poyang Lake of central China from Landsat ETM data. In, *Geoscience and Remote Sensing Symposium, 2004. IGARSS'04. Proceedings. 2004 IEEE International* (pp. 4590-4593): IEEE

Richards, J., & Jia, X. (1999). Remote sensing digital image analysis: An introduction Springer-Verlag. *Berlin Heidelberg*

Richards, J.A. (1996). Classifier performance and map accuracy. *Remote sensing of environment, 57*, 161-166

Richardson, L.L., & LeDrew, E.F. (2006). *Remote sensing of aquatic coastal ecosystem processes*. Springer

Robinson, C., Saatchi, S., Neumann, M., & Gillespie, T. (2013). Impacts of spatial variability on aboveground biomass estimation from L-band radar in a temperate forest. *Remote Sensing, 5*, 1001-1023

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

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C., & Roberts, D. (2008). Mapping land-cover modifications over large areas: A comparison of machine learning algorithms. *Remote sensing of environment, 112*, 2272-2283

Rogers, K., Saintilan, N., & Copeland, C. (2014). Managed retreat of saline coastal wetlands: challenges and opportunities identified from the Hunter River Estuary, Australia. *Estuaries and Coasts, 37*, 67-78

Rosso, P., Ustin, S., & Hastings, A. (2005). Mapping marshland vegetation of San Francisco Bay, California, using hyperspectral data. *International Journal of Remote Sensing, 26*, 5169-5191

Roth, K.L., Roberts, D.A., Dennison, P.E., Peterson, S.H., & Alonzo, M. (2015). The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data. *Remote sensing of environment, 171*, 45-57

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics, 23*, 2507-2517

Sahoo, B.C., Oommen, T., Misra, D., & Newby, G. (2007). Using the one-dimensional S-transform as a discrimination tool in classification of hyperspectral images. *Canadian Journal of Remote Sensing, 33*, 551-560

Salas, C., Ene, L., Gregoire, T.G., Næsset, E., & Gobakken, T. (2010). Modelling tree diameter from airborne laser scanning derived variables: A comparison of spatial statistical models. *Remote sensing of environment, 114*, 1277-1285

Schapire, R.E. (2003). The boosting approach to machine learning: An overview. *Nonlinear estimation and classification* (pp. 149-171): Springer

Scheffler, D., & Karrasch, P. (2013). Preprocessing of hyperspectral images: a comparative study of destriping algorithms for EO1-hyperion. In, *SPIE Remote Sensing* (pp. 88920H-88920H-88915): International Society for Optics and Photonics

Schile, L.M., Callaway, J.C., Morris, J.T., Stralberg, D., Parker, V.T., & Kelly, M. (2014). Modeling tidal marsh distribution with sea-level rise: Evaluating the role of vegetation, sediment, and upland habitat in marsh resiliency. *PloS one, 9*, e88760

Schmidt, K., & Skidmore, A. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote sensing of environment, 85*, 92-108

Schölkopf, B., & Smola, A.J. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press

Schuster, C., Förster, M., & Kleinschmit, B. (2012). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *International Journal of Remote Sensing, 33*, 5583-5599

Sellers, P.J. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing, 6*, 1335-1372

Sesnie, S.E., Finegan, B., Gessler, P.E., Thessler, S., Ramos Bendana, Z., & Smith, A.M. (2010). The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees. *International Journal of Remote Sensing, 31*, 2885-2909

Shaikh, M., Green, D., & Cross, H. (2001). A remote sensing approach to determine environmental flows for wetlands of the Lower Darling River, New South Wales, Australia. *International Journal of Remote Sensing, 22*, 1737-1751

Smith, M.O., Johnson, P.E., & Adams, J.B. (1985). Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis. *Journal of Geophysical Research: Solid Earth, 90*

Smola, A.J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing, 14*, 199-222

Solution, E.V.i. (2016). In

Somdatta, C., & Chakrabarti, S. (2010). Preprocessing of hyperspectral data: a case study of Henry and Lothian Islands in Sunderban Region, West Bengal, India

Song, X., Duan, Z., & Jiang, X. (2012). Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image. *International Journal of Remote Sensing, 33*, 3301-3320

Sridharan, H. (2010). Multi-level urban forest classification using the WorldView-2 8-band hyperspectral imagery. *Digital Globe® 8Bands Research Challenge*, 1-39

Srivastava, P.K., Han, D., Rico-Ramirez, M.A., Bray, M., & Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in space research, 50*, 1250-1265

Stefanski, J., Mack, B., & Waske, B.r. (2013). Optimization of object-based image analysis with random forests for land cover mapping. *IEEE Journal of selected topics in applied earth observations and remote sensing, 6*, 2492-2504

Stehman, S.V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of environment, 62*, 77-89

Stehman, S.V., & Czaplewski, R.L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote sensing of environment, 64*, 331-344

Story, M., & Congalton, R.G. (1986). Accuracy assessment-A user\'s perspective. *Photogrammetric engineering and remote sensing, 52*, 397-399

Strahler, A.H., Boschetti, L., Foody, G.M., Friedl, M.A., Hansen, M.C., Herold, M., Mayaux, P., Morisette, J.T., Stehman, S.V., & Woodcock, C.E. (2006). Global land cover validation: Recommendations for evaluation and accuracy assessment of global land cover maps. *European Communities, Luxembourg, 51*

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007a). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics, 8*, 25

Strobl, C., Boulesteix, A.L., Zeileis, A., & Hothorn, T. (2007b). Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC bioinformatics, 8*

Stumpf, A., & Kerle, N. (2011). Object-oriented mapping of landslides using Random Forests. *Remote sensing of environment, 115*, 2564-2577

Su, L. (2009). Optimizing support vector machine learning for semi-arid vegetation mapping by using clustering analysis. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 407-413

Su, L., & Huang, Y. (2009). Support vector machine (svm) classification: comparison of linkage techniques using a clustering-based method for training data selection. *GIScience & Remote Sensing, 46*, 411-423

Swanson, K.M., Drexler, J.Z., Schoellhamer, D.H., Thorne, K.M., Casazza, M.L., Overton, C.T., Callaway, J.C., & Takekawa, J.Y. (2014). Wetland accretion rate model of ecosystem resilience (WARMER) and its application to habitat sustainability for endangered species in the San Francisco Estuary. *Estuaries and Coasts, 37*, 476-492

Tan, C.-P., Koay, J.-Y., Lim, K.-S., Ewe, H.-T., & Chuah, H.-T. (2007). Classification of multi-temporal SAR images for rice crops using combined entropy decomposition and support vector machine technique. *Progress In Electromagnetics Research, 71*, 19-39

Tanase, M.A., Panciera, R., Lowell, K., Tian, S., Hacker, J.M., & Walker, J.P. (2014). Airborne multi-temporal L-band polarimetric SAR data for biomass estimation in semi-arid forests. *Remote sensing of environment, 145*, 93-104

Tang, D., Jin, W., Qin, N., & Li, H. (2016). Feature selection and analysis of single lateral damper fault based on SVM-RFE with correlation bias reduction. In, *Control Conference (CCC), 2016 35th Chinese* (pp. 3840-3845): IEEE

Tang, Y., Zhang, Y.-Q., & Huang, Z. (2007). Development of two-stage SVM-RFE gene selection strategy for microarray expression data analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 4*, 365-381

Tarantino, C., Adamo, M., Pasquariello, G., Lovergine, F., Blonda, P., & Tomaselli, V. (2012). 8-band image data processing of the worldview-2 satellite in a wide area of applications. *Earth Observation*: InTech

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Legg, C., & De Dieu, M.J. (2004a). Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. *Remote sensing of environment, 90*, 23-43

Thenkabail, P.S., Enclona, E.A., Ashton, M.S., & Van Der Meer, B. (2004b). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote sensing of environment, 91*, 354-376

Thenkabail, P.S., Mariotto, I., Gumma, M.K., Middleton, E.M., Landis, D.R., & Huemmrich, K.F. (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*, 427-439

Thenkabail, P.S., Smith, R.B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote sensing of environment, 71*, 158-182

Thomson, A., Fuller, R., Yates, M., Brown, S., Cox, R., & Wadsworth, R. (2003). The use of airborne remote sensing for extensive mapping of intertidal sediments and saltmarshes in eastern England. *International Journal of Remote Sensing, 24*, 2717-2737

Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random Forest Classification of Wetland Landcovers from Multi-Sensor Data in the Arid Region of Xinjiang, China. *Remote Sensing, 8*, 954

Tso, B.C., & Mather, P.M. (1999). Classification of multisource remote sensing imagery using a genetic algorithm and Markov random fields. *Geoscience and Remote Sensing, IEEE Transactions on, 37*, 1255-1260

Tucker, C.J. (1977). Asymptotic nature of grass canopy spectral reflectance. *Applied Optics, 16*, 1151-1156

Tuia, D., Pacifici, F., Kanevski, M., & Emery, W.J. (2009). Classification of very high spatial resolution imagery using mathematical morphology and support vector machines. *IEEE Transactions on Geoscience and Remote Sensing, 47*, 3866-3879

Uhlmann, S., & Kiranyaz, S. (2014). Integrating color features in polarimetric SAR image classification. *IEEE Transactions on Geoscience and Remote Sensing, 52*, 2197-2216

van Beijma, S., Comber, A., & Lamb, A. (2014). Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote sensing of environment, 149*, 118-129

Vapnik, V.N., & Kotz, S. (1982). *Estimation of dependences based on empirical data*. Springer-Verlag New York

Wang, H., Zhao, Y., Pu, R., & Zhang, Z. (2015). Mapping Robinia pseudoacacia forest health conditions by using combined spectral, spatial, and textural information extracted from IKONOS imagery and random forest classifier. *Remote Sensing, 7*, 9020-9044

Wang, L.a., Zhou, X., Zhu, X., Dong, Z., & Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop Journal, 4*, 212-219

Wang, Y., Traber, M., Milstead, B., & Stevens, S. (2007). Terrestrial and submerged aquatic vegetation mapping in Fire Island National Seashore using high spatial resolution remote sensing data. *Marine Geodesy, 30*, 77-95

Waske, B., & Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 450-457

Weinmann, M., Jutzi, B., Hinz, S., & Mallet, C. (2015). Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 286-304

Weston, J., Elisseeff, A., Schölkopf, B., & Tipping, M. (2003). Use of the zero-norm with linear models and kernel methods. *Journal of machine learning research, 3*, 1439-1461

Winning, G., & Saintilan, N. (2009). Vegetation changes in Hexham Swamp, Hunter River, New South Wales, since the construction of floodgates in 1971. *Cunninghamia, 11*, 185-194

Wolf, A. (2010). Using WorldView 2 Vis-NIR MSI imagery to support land mapping and feature extraction using normalized difference index ratios. In, *Proc. SPIE*

Xiao, J., & Moody, A. (2005). A comparison of methods for estimating fractional green vegetation cover within a desert-to-upland transition zone in central New Mexico, USA. *Remote sensing of environment, 98*, 237-250

Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical, 212*, 353-363

Yang, J., Ren, G., Ma, Y., & Fan, Y. (2016). Coastal wetland classification based on high resolution SAR and optical image fusion. In, *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International* (pp. 886-889): IEEE

Yokoya, N., Miyamura, N., & Iwasaki, A. (2010). Preprocessing of hyperspectral imagery with consideration of smile and keystone properties. In, *SPIE Asia-Pacific Remote Sensing* (pp. 78570B-78570B-78579): International Society for Optics and Photonics

Yu, L., Fu, H., Wu, B., Clinton, N., & Gong, P. (2016). Exploring the potential role of feature selection in global land-cover mapping. *International Journal of Remote Sensing, 37*, 5491-5504

Yuan, J., & Niu, Z. (2008). Evaluation of atmospheric correction using FLAASH. In, *Earth Observation and Remote Sensing Applications, 2008. EORSA 2008. International Workshop on* (pp. 1-6): IEEE

Zhang, C., Wang, T., Atkinson, P.M., Pan, X., & Li, H. (2015). A novel multi-parameter support vector machine for image classification. *International Journal of Remote Sensing, 36*, 1890-1906

Zhang, M., Gong, Z., Zhao, W., Pu, R., & Liu, K. (2016). Estimating wetland vegetation abundance from Landsat-8 operational land imager imagery: a comparison between linear spectral mixture analysis and multinomial logit modeling methods. *Journal of Applied Remote Sensing, 10*, 015005-015005

Zhang, R., & Ma, J. (2009). Feature selection for hyperspectral data based on recursive support vector machines. *International Journal of Remote Sensing, 30*, 3669-3677

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote sensing of environment, 113*, 182-196

Zhu, G., & Blumberg, D.G. (2002). Classification using ASTER data and SVM algorithms;: The case study of Beer Sheva, Israel. *Remote sensing of environment, 80*, 233-240

Zomer, R.J., Trabucco, A., & Ustin, S.L. (2009). Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing. *Journal of Environmental Management, 90*, 2170-2177