Using remotely piloted aircraft imagery to quantify faunal impacts on Dirk Hartog Island following ungulate removal

Lucy Wilson / l920-3869 / lucylucywilson@gmail.com

This research report is presented for the Masters of Geospatial Science,
Curtin University

November, 2019

Abstract

Over the last 10 years, Dirk Hartog Island (DHI), located off the coast of Western Australia, has ceased being a pastoral lease and re-established as a national park. Introduced feral animals such as goats and cats have been eradicated as goats are notorious for land degradation and cats feed on small mammals. The DBCA subsequently introduced two hare-wallabies (rufous and banded) to the island in 2017, which are extinct on mainland Australia. Rufous and banded hare-wallabies are considered ecosystem engineers as their digging behaviour creates a physical state change to the environment and modulates resource availability to other organisms. This report aims to provide a base-line dataset for long-term monitoring of the translocated fauna.

To this end, very high-resolution (c. 5mm) Remotely Piloted Aircraft (RPA) imagery was collected with a DJI Phantom 4 over two exclusion plots. Exclusion plots were constructed to protect areas of vegetation from ground-dwelling fauna. RPA data was classified to a species level using a stratified, systematic random sampling of vegetation. Object-based modelling using a machine learning algorithm was employed to classify species. Scale parameterisation of segments was tested for 20, 50, and 100 values. This measure controls the segment object size. Segmentation of RPA imagery derived 21 variables comprising spectral reflectance, green leaf algorithm measures, textures, and shape metrics.

A novel dimension reduction technique was implemented using box plot analysis and Multi-variate Analysis of Variance (MANOVA) assumption testing. Classification scenarios tested were a scale of 20 with all variables, scale of 50 with untransformed selected variables, scale of 50 with selected variables transformed to fit the normal distribution, scale of 50 with all variables, and scale of 100 with all variables. The Random Forest machine learning algorithm was tested for 10 iterations per scenario to ensure stability of predicted classes with mean accuracy and Kappa values recorded for the final result. A 15 x 15 m quadrat was estimated within and external to the constructed exclusion plots on DHI for the purpose of comparing vegetation cover.

The MANOVA and each of the final 5 individual subset variables per a post-hoc univariate test were found to be significant ($p<0.05$) for both exclusion plot sites. Random Forest out-of-bag estimates showed accuracy for x01 to be 75%, 74%, 69%, and 59% for the 20 - all variables, 50 - all variables, 100 - all variables, and 50 - subset variables respectively. Out-of-bag accuracy calculation for

x06 were 78%, 74%, 66%, and 63% for the 20 - all variables, 50 - all variables, 100 - all variables, and 50 - untransformed subset variables respectively. Dimension reduction was shown to improve computational performance by 63.64% and 68.18% for x01 and x06 respectively. Species percentage cover showed a negligible change in vegetation for both exclusion plot sites. This preliminary investigation has shown that RPA imagery derived variables may be used to monitor plant species and small scale faunal impacts using object-based classification techniques.

Acknowledgments

My sincerest gratitude to **Dr. Todd Robinson** of Curtin University for his unwavering support and guidance whilst writing this Master's thesis. He consistently made himself available when I needed help and provided sound feedback with directions to improve. The knowledgeable responses he provided assisted in shaping the statistical analysis and facilitated an in depth inquiry into extracting meaningful information from real-world data.

To **Ricky Van Dongen** from the Department of Biodiversity, Conservation and Attractions, Government of Western Australia for his collaborative assistance throughout this research process. His expertise for coding highly technical algorithms and processing large datasets was instrumental to deriving a species level classification of remotely piloted aircraft imagery. I could not reiterate enough how much I appreciate Ricky's support in helping me to accomplish this Master's thesis.

I would like to recognise the monumental effort of **Katherine Zdunic** from the Department of Biodiversity, Conservation and Attractions, Government of Western Australia during the Dirk Hartog Island field survey component of this study. Her assistance in sampling vegetation enabled us to achieve a complete documentation of species within a restricted timeframe.

Finally, thank you to **Paul Rampant** from the Department of Biodiversity, Conservation and Attractions, Government of Western Australia for capturing the September, 2018 remotely piloted aerial imagery analysed in this report.

Table of contents

List of figures

List of tables

1. Introduction

1.1. Ecosystem engineers, the object-based paradigm, and machine learning

Detrimental human impacts on the environment are a global phenomenon (Ceballos et al., 2015; Ceballos, Ehrlich, & Dirzo, 2017; Ehrlich & Ehrlich, 2013; Hooper et al., 2012; Turney et al., 2008; Vitousek, Mooney, Lubchenco, & Melillo, 1997). One example is the transference of a population into a novel habitat (Garnett, Latch, Lindenmayer, & Woinarksi, 2018). Historically, many Australian species introductions have been unintended or implemented for conservation, however the flow on consequences for the ecosystem they are introduced into has often been deleterious (Campbell et al., 2016; Clarke, Crossland, & Shine, 2016; Hobbs, 1993; Hobbs & Hopkins, 1990; Iannella, Peacock, Cassey, & Schwensow, 2019; Matthew L Brooks et al., 2004; Pittock, Finlayson, & Howitt, 2013; Russo et al., 2018; Taylor & Goldingay, 2010). These impacts are usually due to a limited understanding around how the ecosystem functions holistically (Lindenmayer, 2007). Future threatened species management plans may be improved if we are informed of the interplay between individual behaviours and ecosystem function.

Anthropogenic global warming is likely to strain environmental restoration efforts via climatic stressors such as water scarcity and an increase in extreme weather events (Hoegh-Guldberg et al., 2018). Protecting our endangered species through captive breeding programs, translocations and other conservation efforts is fundamental for the survival of our endemic landscape (Molles, 2009). A thorough understanding of ecosystem processes and long-term monitoring programs are required to ensure conservation efforts are efficacious.

Dirk Hartog Island (DHI), Western Australia was reclaimed from pastoralists by the Department of Biodiversity, Conservation and Attraction and established as a national park in 2009 (Asher & Morris, 2015). The purpose of converting the area into a national park was to create a refuge for native Australian fauna that are extinct on mainland Australia (Asher & Morris, 2015). Mechanisms for Australian species population decline include alien species competition, hard-hoof trampling, and introduced predators (Garnett et al., 2018; Mulligan, Buxton, Lane, Neave, & Richardson, 2015). Feral goats, sheep, and cats were recognised as potential threats to biodiversity on DHI (Asher & Morris, 2015). Exotic ungulates (sheep and goats) were completely removed or eradicated from DHI in 2017 as were feral cats in 2018 (Heriot, Asher, Williams, & Moro, 2019). The removal of invasive fauna species is suggested to alleviate predation and completion pressures to native macropods, which may

impact the success of introducing fauna for conservation purposes (Bode et al., 2013; Hilmer, Algar, & Johnston, 2010).

Banded (*Lagostrophus fasciatus*) and rufous (*Lagorchestes hirsutus*) hare-wallaby populations were translocated to DHI in 2017 (DBCA, 2017). These species are extinct on mainland Australia. The banded hare wallaby is known to create nests within dense bush thickets, particularly in association with *Acacia ligulata*, commonly known as the umbrella bush (Burbidge & Woinarski, 2016b; Cowen, Rayner, Sims, & Morris, 2018). Similarly, the rufous hare-wallaby shelters and digs small burrows within dense thickets/spinifex hummocks such as *Acacia ligulata* and *Triodia plurinervata* (Burbidge & Woinarski, 2016a; Cowen et al., 2018). The digging behaviours of both species may alter the structure of coastal dune ecosystems found on DHI, which could result in the modulation of resources to locally established taxa (Munro et al., 2019; Valentine et al., 2016). For example burrows may create habitat or refugia for other organisms or alter the soil chemistry and increase leaching to promote plant growth (Crooks, 2002). This process is known as ecosystem engineering (Jones, Lawton, & Shachak, 1994). Both of the translocated endangered species to DHI have been recognised as ecosystem engineers (Manning, Eldridge, & Jones, 2015). However, there is an absence of empirical evidence on the potential influences ecosystem engineers have on coastal landscapes.

Burrowing impacts of faunal ecosystem engineers on vegetation cover has shown an increase in both animal and plant species heterogeneity (Louw, Roux, Meyer-Milne, & Haussmann, 2017; Streitberger & Fartmann, 2016). However, these estimates are based on gradual distances from known burrowing sites. There are limited examples within the literature which use a discrete exclusion zone to minimise the potential for stochastic species migrations to confound the outcome. Furthermore, there is also a need for studies calculating the flow-on effects digging mammals have on overall vegetation cover. These findings may allow scientist to better understand the broader implications of introducing a species into a novel habitat for conservation purposes.

Further benefits of faunal digging processes include reducing soil hydrophobicity (Eldridge & Mensinga, 2007; Valentine et al., 2017); increased soil nutrient content (Eldridge & James, 2009; James, Eldridge, & Hill, 2009; Mallen-Cooper, Nakagawa, & Eldridge, 2019; Travers, Eldridge, Koen, & Soliveres, 2012; Valentine et al., 2018); improve soil density (Cuevas, Mastrantonio, Ojeda, & Jaksic, 2012; Travers et al., 2012); restore species heterogeneity (Grossman, Hayward, & Gibb, 2019; Louw

et al., 2017; Streitberger & Fartmann, 2016); and influence seed germination through the creation of microhabitats (James et al., 2009; Sandom, Hughes, & Macdonald, 2013; Valentine et al., 2016). Ecosystem engineering by digging fauna may also improve plant establishment and development. Valentine et al. (2018) found that biopedturbation of the quenda bandicoot (*Isoodon fuscicenter*) lead to an increase in plant growth. However, studies measuring digging impacts on species cover and richness have predominately focused on localised sampling from known burrowing sites (Louw et al., 2017; Streitberger & Fartmann, 2016). Research on species cover and richness using remotely sensed imagery may improve conservation recovery efforts through informed decisions of the flow-on effects across a wider observation area.

Remotely sensed aerial imagery provides a valuable data source for earth observation and geoinformatics (Belward & Skøien, 2015; Kelly, Blanchard, Kersten, & Koy, 2011; Lillesand, Kiefer, & Chipman, 2015; Melesse, Weng, Thenkabail, & Senay, 2007). Assessment of vegetation may be conducted using imagery acquired from a satellite or Remotely Piloted Aircraft (RPA) also referred to as an Unmanned Aerial Vehicle or UAV (Ali, Qazi, & Aslam, 2018; Bolyn, Michez, Gaucher, Lejeune, & Bonnet, 2018; Dronova, 2015; Ghosh, Fassnacht, Joshi, & Koch, 2014; Ochoa & Guo, 2019; Silveira et al., 2019; Waldner et al., 2016; Zhang, Denka, Cooper, & Mishra, 2018). Advances in remote sensing technology has enabled scientists to capture sub-centimeter aerial imagery (Ochoa & Guo, 2019). The development of aerial imagery capture allows for classification to be made across a range of scales between land use and land cover to delineating tree species (Belgiu & Dra˘gut, 2016; Chen et al., 2018; McInerney, Kempeneers, Marron, & McRoberts, 2019; Wu, Zhong, Zhao, Fu, & Song, 2017; Zhou et al., 2018). These studies require different spatial resolutions depending on what is an appropriate scale level for the class types being assessed (Lu, Hetrick, & Moran, 2011; Phillips, 2013).

Resolution is dependent on the distance between the instrument capturing the aerial imagery and the ground (Tang & Shao, 2015). Aerial imagery on DHI is estimated to have a spatial resolution of between 5-6mm based on a pre-recorded flight plan. This was the selected method of capture as measuring vegetation to a species level is a fine-scale study requiring detailed aerial imagery. There have been limited studies using sub-centimeter spatial resolution (Ochoa & Guo, 2019). Furthermore, there are no papers quantifying coastal heathland vegetation to a species level using object-based methods. Determining an accurate delineation of the endemic flora mosaic found on DHI using fine-scale imagery is valuable for shaping future vegetation monitoring approaches.

Traditional methods of classifying aerial imagery focus on the spectral reflectance per pixel (Bolyn et al., 2018; Kumar, Dasgupta, Mukhopadhyay, & Ramachandra, 2018; Zhang, Xiao, & Feng, 2012). Pixel based approaches may be limited when measuring imagery with high spectral variability within classes as found in RPA capture (Dronova, 2015; Lu & Weng, 2007; Zhang, 2016; Zhou et al., 2018). Furthermore, contextual information such as texture and pixel adjacency is not factored into the classification output (Ali et al., 2018; Blaschke, 2010; Townshend et al., 2000). Inclusion of these variables through an object-based imagery analysis may improve the classification outcome.

An object-based image analysis groups pixels into homogenous segments (Kim, Madden, & Warner, 2009; Pu & Landry, 2012; Sherba, Blesius, & Davis, 2014). These segments are also commonly referred to as image objects and may be used in aerial imagery classification (Ye, Pontius, & Rakshit, 2018). Using image objects mitigates issues arising from individual pixel classification techniques as it allows for geospatial variables to be measured, pixel adjacency considerations, and spectral heterogeneity to be encapsulated into polygon areas (Benz, Hofmann, Willhauck, Lingenfelder, & Heynen, 2004; Kawakubo, Morato, & Luchiari, 2013; Silveira et al., 2019; Zhang et al., 2012). Another benefit to an object-based imagery analysis is that uncertainties arising from positional discrepancies are likely to be reduced (Congalton & Green, 2009). Plant samples have a higher probability of being located within a segment then a pixel (Addink, Jong, & Pebesma, 2007). Therefore, spatial accuracy of the classification may also be improved.

Machine learning techniques are rapidly emerging as an effective method to accurately classify big imagery datasets (Chapman, Bonn, Kunin, & Cornell, 2010; Jordan & Mitchell, 2015; Teluguntla et al., 2018). It is a method of automating data analysis based on the computer assimilating and adapting from data, recognising patterns, anticipating responses, and making strategic decisions (Mitchell, 1997).

Random Forest (RF) is a widely used machine learning algorithm as the technique is insusceptible to overfitting and may compute multidimensional variables (Chen et al., 2018). Initially proposed by Breiman (2001), RF is a non-parametric algorithm which utilises a 'forest' of decision trees to predict a surface based on individually trained classes. The strength of the outcome is derived from the ability for RF to apply more than one classifier, it does not prune the data, and randomly

selects training samples to split nodes (Breiman, 2001; Cutler, Edwards, Beard, & Cutler, 2007; Duro, Franklin, & Dubé, 2012; Ho, 1998). Therefore, it is expected that RF will be suitable as a long-term method for classifying remotely sensed aerial imagery at a species level.

1.2. Problem statement

Australia has experienced strong negative effects from our efforts to restore declining environmental conditions (Campbell et al., 2016; Clarke et al., 2016; Hobbs, 1993; Hobbs & Hopkins, 1990; Iannella et al., 2019; Matthew L Brooks et al., 2004; Pittock et al., 2013; Russo et al., 2018; Taylor & Goldingay, 2010). It is evident that our efforts to preserve the endemic Australian landscape may be improved through informed decisions. This is especially pertinent when considering the anthropogenic introduction of animals into a novel habit (Garnett et al., 2018). Ecosystem engineering is a relatively underdeveloped area of biological study (Jones et al., 2010). There are a myriad of unknowns and lurking variables surrounding the consequences of fauna modulating resources for other organisms. Furthermore, there has been limited studies quantifying the impacts digging behaviours have on the broader distribution of vegetation cover. Understanding how translocating species impact landscapes via ecosystem engineering may improve environmental conservation outcomes.

Determining whether ecosystem engineers broadly affect vegetation requires sound species level data capture and classification methods. Especially when considering the refined spatial resolution RPA imagery captured on DHI. However, there is limited research utilising sub-centimeter remotely sensed imagery. The broader theory has shown that traditional per-pixel approaches are inept when dealing with high-resolution, multi-model datasets which are known to contain increased spectral heterogeneity (Belgiu & Csillik, 2018; Jebur, Shafri, Pradhan, & Tehrany, 2014; Lu, Hetrick, & Moran, 2010; Ma et al., 2017; Malahlela, Cho, & Mutanga, 2014; Myint, Gober, Brazel, Grossman-Clarke, & Weng, 2011; Niphadkar, Nagendra, Tarantino, Adamo, & Blonda, 2017; Radoux, Bogaert, Fasbender, & Defourny, 2011; Silveira et al., 2019; Wu et al., 2017; Zhang et al., 2018). Object-based approaches which apply a machine learning classifier may handle the complexities embedded in fine scale data. (Adam, Mutanga, Odindi, & Abdel-Rahman, 2014; Du, Samat, Waske, Liu, & Li, 2015; Jhonnerie, Siregar, Nababan, Prasetyo, & Wouthuyzen, 2015; Noi & Kappas, 2017; Pantaleoni, Wynne, Galbraith, & Campbell, 2009; Raczko & Zagajewski, 2017; Rodriguez-Galiano, Chica-Olmo, Abarca-Hernandez, Atkinson, & Jeganathan, 2012). Within the mosaic of sub-centimeter studies, there

are none which classify coastal dune communities to a species level. Delineating viable methods for classifying RPA capture to a species level is important in determining effective long-term vegetation monitoring.

1.3. Study objectives

The aim of this research is to implement an object-based approach using remotely piloted aircraft imagery to measure the ecosystem engineered impacts of translocated fauna on DHI. Five objectives have been developed to achieve this aim: a) construct and test a framework of sampling vegetation for the purpose of classifying high-resolution aerial imagery b) segmentation of aerial imagery and implementation of the RF algorithm to classify segmented data, c) dimension reduction to obtain variables offering the greatest degree of species level separation, d) accuracy assessment of the RF classifier, and e) determining if translocated fauna are having an ecosystem engineered impact on DHI. A comparison of vegetation cover between areas within and external to the 40 x 40 m exclusion plots may determine if the introduced burrowing fauna are influencing plant growth and/or species richness. Understanding the flow on effects translocating digging fauna may have on vegetation could improve the success of future attempts to bio-remedy the environment and preserve critically endangered species. The methods delineated in this paper may provide a long-term vegetation monitoring framework for conservation managers.

1.4. Thesis structure

This thesis consists of nine chapters in total. Chapter one introduces the importance of the study including the current global trend of environmental degradation; the history of DHI; the potential for the introduced fauna to act as ecosystem engineers; advances in remotely sensed data capture; techniques to classify aerial imagery; and the aims and objectives of this paper.

Secondly, a comprehensive review of the scientific literature will be conducted in chapter two. Research gaps within the available scientific literature will be identified. The chapter will begin with a detailed explanation of ecosystem engineering and the specific benefits of digging behaviours. The inference that ecosystem engineering behaviours may correlate with differences in percentage cover between exclusion plots and proximal areas will be supported per the breadth of studies measuring direct engineering impacts. The importance of developing ecosystem engineering theory will be defined. Methods to quantify remotely sensed data capture to a species level for long-term vegetation

monitoring may be guided per analysis of previous object-based approaches. Studies implementing segment parameterisation comparisons, machine learning approaches, and auxiliary variables to classify remotely sensed imagery will be scrutinized in order to facilitate an accurate depiction of the mosaic of vegetation on DHI.

Chapter three comprises the methods and materials applied within this research paper. The characteristics for both exclusion plot study areas will be provided along with the means of sampling plant and aerial imagery within the study sites. This chapter will also detail the process of segmenting remotely piloted aircraft data, statistically quantifying pertinent variables which will allow for the measurement of vegetation to a species level, and the application of RF to classify the remotely sensed dataset. Lastly the technique to estimate differences in vegetation cover between the exclusion plot and proximal areas will be outlined.

Results detailed in chapter four will directly align with the methods from the previous chapter. This study will address the results from the vegetation samples identified, segmentation parameters, descriptive and inferential statistics determining remotely sensed variables to delineate plant species, the RF classification output, an accuracy assessment error matrix, and the comparisons of percentage vegetation cover.

The discussion may be found in chapter five and concluding remarks and recommendations for further studies will be provided in chapter six. Discussion topics include survey methodologies and improved sampling for future vegetation monitoring programs; reducing the 'curse of dimensionality'; segment parameterisation; alternative independent accuracy assessments e.g. a confusion matrix; and small scale faunal impacts on DHI vegetation.

Lastly, chapter seven shows the bibliography which outlines the references cited within this thesis and chapter eight is the appendix. The appendices in this instance details the R code implemented within the methods section and supplementary data to the results section. The R code may be replicated for future studies which require statistical analytics of segmented data and classification using the RF machine learning algorithm.

2. Scientific literature review

This chapter critically analyses the current body of scientific literature surrounding the following topics: 2.1) Ecosystem Engineers, 2.2) comparisons of per-pixel and object-based classification, 2.3) segmenting Remotely Piloted Aircraft (RPA) capture, 2.4) classifying segments using methods derived for per-pixel approaches, 2.5) machine learning algorithms, 2.6) additional variables that may compliment spectral reflectance for species level classification, and 2.7) measuring the engineering impacts of digging fauna. An understanding of what we currently know will assist in determining the best method for a species level classification of RPA imagery. Research gaps for the aforementioned topics are identified. Disadvantages of per-pixel based methods for classifying aerial imagery will be compared with studies incorporating an object-based approach.

2.1. Ecosystem engineering

Ecosystem engineering processes are characterised by species whom directly or indirectly create a physical state change within their environment, resulting in the modulation of resources to other organisms (Figure 2.1)(Jones et al., 1994). Modulation in this instance does not refer to the provision of edible carrion, leaves, fruits, and other food resources as this would constitute a trophic interaction (Linder et al., 2012). An ecosystem engineer alters biotic and abiotic materials for the subsequent modification, creation, maintenance, and/or destruction of habitat (Jones, Lawton, & Shachak, 1997). Recognizing the consequences of these impacts and their involvement within complex interaction webs may enable a greater understanding of ecosystem function in its entirety (Eisenhauer, Milcu, Sabais, & Scheu, 2008).

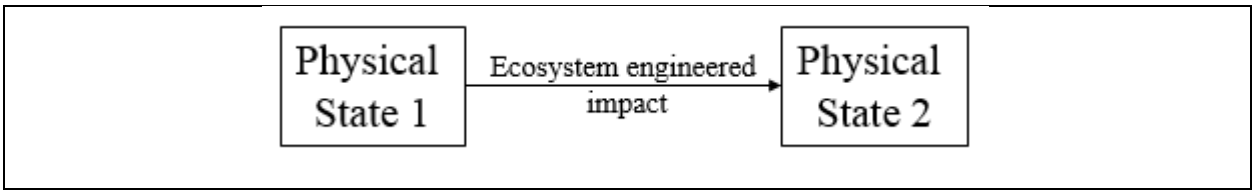

Figure 2.1. Conceptual diagram showing the process of a physical state change within an environment caused by an ecosystem engineer.

In Western Australia, *Leptospermum laevigatum* has naturalized beyond its native range (Queensland Government, 2016). Whilst locally invasive, the species is recognised for the ecosystem services provided in stabilising coastal dune systems and buffeting sea spray (Paczkowska, 1995b). Effective engineering of harsh coastal environments as a direct consequence of the taxon's physical

impact allows for the colonisation of further plant life (Reubens, Poesen, Danjon, Geudens, & Muys, 2007). This habitat creation constitutes an ecosystem engineered process. Other commonly cited examples of ecosystem engineering include beavers (*Castor Canadensis*) constructing dams thus altering the local hydrology, which effects the composition and diversity of aquatic inhabitants (Pollock et al., 1995); earthworm seed digestion functioning as a secondary disperser (Eisenhauer et al., 2008); the creation of micro-habitats from flora debris (Hedin, Mayer, & Likens, 1988); and the role of microbial communities in engineering soil crusts (Xiao, Sun, Hu, & Kidrond, 2019).

At some level every organism may engineer their environment (Jones et al., 1994). The ubiquity of impacts across all biomes and environmental scales is a fundamental reason why ecosystem engineering has historically been overlooked by environmental scientists (Jones et al., 2010). Ecosystem engineering theory has been absent from biological text-books which traditionally focus on well-established concepts (population dynamics, species interactions, niche adaptations, nutrient cycles, succession and stability, etc.) (Levin, 2009; Molles, 2009). Therefore, it is imperative that there is a research focus to develop the body of literature comprising the theory.

Faunal digging has been established as a method of ecosystem engineering (Manning et al., 2015; Munro et al., 2019). Digging processes improve water filtration (Eldridge & Mensinga, 2007; Valentine et al., 2017); increase nutrients and/or nutrient cycling (Eldridge & James, 2009; James et al., 2009; Mallen-Cooper et al., 2019; Travers et al., 2012); reduce soil bulk-density (Cuevas et al., 2012; Travers et al., 2012); restore landscape heterogeneity (Grossman et al., 2019; Louw et al., 2017; Streitberger & Fartmann, 2016); and create microhabitats for seed germination (James et al., 2009; Sandom et al., 2013; Valentine et al., 2016). Recognising these ecosystem engineered impacts may improve conservation efforts. Munro et al. (2019) found that returning the locally extinct eastern bettong (*Bettongia gaimardi*) to Mulligans Flat Woodland Sanctuary, Canberra had a significant impact on the soil composition. Biopedturbation (animal driven soil disturbance) by the eastern bettong was predicted at a rate of 985 kg of soil per hectare annually. This was significantly higher than comparative soil turnover rates of the rabbit, echidna and bird samples at 622, 159, and 66kg/ha/year respectively (Munro et al., 2019). This shows that Australian digging fauna have traditionally provided a greater ecological service than exotic species and non-digging animals in tilling soils.

The intrinsic characteristics of faunal derived pits may also affect the delivery of ecosystem services. Ross et al. (2019) compared eastern bettong and rabbit pits in relation to their morphology, longevity, temperature, and nutrients. They found eastern bettong pit dimensions to be narrower and deeper than rabbits, persist in the environment for over 2 years, moderate diurnal temperatures by up to 25°C, and contain greater carbon and nitrogen amounts than a control site as a result of increased organic litter collection (Ross et al., 2019). However, many digging mammals have experienced severe range contractions and population declines predominately due to habitat loss/fragmentation, disease, and introduced predators, resulting in a loss of the ecosystem services they provide (Davidson, Detling, & Brown, 2012; Fleming et al., 2014; Woinarski, Burbidge, & Harrison, 2015). The translocation of native species may restore lost ecological functions via ecosystem engineering.

Verdon, Gibb, and Leonard (2016) studied the role of the bridled nailtail wallaby, greater bilby, brush-tailed bettong, burrowing bettong, numbat, and greater stick-nest rat have on restoring ecosystem function in Scotia Sanctuary, New South Wales. They found that exclusion plots had a higher abundance of naturally established seedlings (MEAN 4 ± 1.11SE), subshrubs (MEAN 4.03 ± 0.08SE) and perennial forbs (MEAN 1.96 ± 0.27SE) compared with a control plot (MEAN and SE of 1.5 ± 0.5, 3.55 ± 0.11, and 1.22 ± 0.15 respectively) (Verdon, Gibb, & Leonard, 2016). However, they recognise that these differences may be the result of both soil disturbance and herbivory.

Valentine et al. (2018) examined whether biopedturbation of the quenda bandicoot (*Isoodon fuscicenter*) due to foraging behaviours promoted plant growth in foraging pits, associated spoil heaps, and a control area. Measurements were solely on the pit characteristics and therefore herbivory was removed as a potential confounding variable. They found that the spoil heap had the greatest level of conductivity, phosphorus, potassium, sulphur, and microbial activity when compared with foraging pits and control areas (Valentine et al., 2018). Sub-sequent seedling growth from the spoils heaps were 1.5 to 2 times faster than the control or foraging pit. Test samples were taken from fresh pits rather than those which had been extant for elongated periods of time. Foraging pits may experience inertia when providing beneficial growing conditions through organic litter build up and altered abiotic conditions for microhabitat creation as found in other studies (Eldridge & James, 2009; James et al., 2009; Mallen-Cooper et al., 2019; Sandom et al., 2013; Travers et al., 2012; Valentine et al., 2016). Overall the research found that quenda digging behaviours are conducive to promoting plant growth.

Digging behaviours show an have shown to increase both animal and plant species heterogeneity from known burrowing sites (Louw et al., 2017; Streitberger & Fartmann, 2016). However, evidence-based research implementing exclusion plot controls to reduce bias are limited. There is also a literature gap for studies on how soil biopedturbation may affect overall vegetation cover. These findings may allow scientist to better understand the broader implications of introducing a species to a novel habitat for conservation purposes.

2.2. Per-pixel vs. object-based image classification

RPA data capture results in the earth's surface being represented as a raster layer (Lillesand et al., 2015; Melesse et al., 2007). Raster images are comprised of individual pixels that denote the spatial resolution and ascribe a radiance value in the form of a Digital Number (DN) (Pavlidis, 2017; Richards, 2013). Classification of aerial imagery into feature types may be orchestrated by a per-pixel or object-based approach (Arbiol, Zhang, & Palà, 2006; Ma et al., 2017). Common per-pixel methods of classification incorporate a moving window or kernal to average the surrounding DN values (Chavez & Bauer, 1982).

Maximum-Likelihood (ML) is a frequently implemented classification algorithm which assigns classes based on the variance and co-variance of signatures or training samples (Lillesand et al., 2015). Parametric classifiers such as ML rely on distributional referencing when modelling training groups (McInerney et al., 2019). ML may have difficulty classifying ground surface features that are not homogenous (Bischof, Schneider, & Pinz, 1992). Where the spread of training sample classes are not Gaussian, the ML classification technique may not accurately represent surface features (Ali et al., 2018). Therefore, the distinction between the ground surface features is diminished where spatial heterogeneity is high and may contribute to 'salt and pepper' noise.

High-resolution capture classified by applying a per-pixel approach may result in the 'salt and pepper' phenomena (Lu et al., 2011). When the instantenous field of view of the sensor is smaller than the size of the object being record, or the object is highly heterogenous, the classifier may not be able to determine the accurate class (Kelly et al., 2011). In this instance the value assigned to the pixel is either indeterminate or incorrect. The level of detail found in fine-scale datasets may contribute to 'salt and pepper' effects as the pixel spectral heterogeneity cannot be accounted for (Zhang, 2016).

Inaccuracies resulting from high-resolution aerial imagery per-pixel classification may be mitigated using an object-based paradigm (Blaschke, 2010; Dronova, 2015; Zhou et al., 2018).

Silveira et al. (2019) compared object-based modelling using Random Forest (RF) to a pixel based approach. Their methodology measured Landsat-5 TM spectral reflectance, Shuttle Radar Topography Mission digital elevation model, and bioclimatic variables. Results show that the above-ground forest biomass calculations using RF returned a lower mean absolute error (20.95%) than the pixel-based unit (28.64%) (Silveira et al., 2019). Lu, Hetrick, and Moran (2011) also found that when comparing classified segmented imagery to a pixel based approach, an object-based approach reduced speckle resulting from the 'salt and pepper' phenomena. An object-based study of the fine-scale RPA capture on Dirk Hartog Island (DHI) may reduce probably pixelated noise and provide a reliable outcome for long-term vegetation monitoring.

The object-based paradigm takes raster imagery pixels and clusters them into homogenous areas. (Kim et al., 2009; Pu & Landry, 2012; Sherba et al., 2014). These object areas or segments are used as the unit for classifying remotely sensed data (Ye et al., 2018). After segmenting a remotely sensed dataset, the final output is a contiguous layer of object areas (Benz et al., 2004). Image objects reduces noise arising from the 'salt and pepper' phenomenon (Blaschke, 2010; Dronova, 2015; Silveira et al., 2019; Zhang, 2016; Zhou et al., 2018). Another advantage is that segmented regions may hold additional information such as shape and texture which are not accessible in a per-pixel approach (Chen et al., 2018; Niphadkar et al., 2017). An object-based image analysis may also improve classification through pixel adjacency considerations and the ability to encapsulate spatial heterogeneity into polygon areas (Benz et al., 2004; Kawakubo et al., 2013; Silveira et al., 2019; Wu et al., 2017; Zhang et al., 2012).

Comparisons between the two paradigms have repeatedly shown that the object-based method outperforms a per-pixel analysis (Belgiu & Csillik, 2018; Jebur et al., 2014; Lu et al., 2010; Ma et al., 2017; Malahlela et al., 2014; Myint et al., 2011; Niphadkar et al., 2017; Radoux et al., 2011; Silveira et al., 2019; Wu et al., 2017; Zhang et al., 2018). However, more rarely is a higher accuracy achieved from the per-pixel methods (Adam, Csaplovics, & Elhaja, 2016; Duro et al., 2012). Duro, Franklin, and Dube (2012) used a McNemar statistical test to compare pixel and object-based approaches using the same decision tree classification method on the South Saskatchewan River, Canada. They found

that there was no statistical difference between the accuracy of the per-pixel and object-based decision tree classifiers ($p > 0.05$). However, when they compared the object-based decision tree result with Support Vector Machine (SVM) and RF, both achieved a significant difference ($p < 0.01$). This suggests that decision tree used to compare the per-pixel and object-based approaches may have been ill-suited to quantifying segmented imagery.

A similar experiment to Duro, Franklin and Dube (2012) using the McNemar statistical test to compare per-pixel and object-based image classification was conducted by Yan et al. (2007). They classified potential surface coal fire areas in Wuda, China from ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) satellite imagery (Yan, Mas, Maathuis, Xiangmin, & Dijk, 2007). There was a statistical significance between the per-pixel and object-based classification with the latter returning a 36.77% increase in overall accuracy.

Lu, Hetrick, and Moran (2010) compared a per-pixel classification to segmentation using the ML method and an object-based approach called Extraction and Classification of Homogenous Objects (ECHO) partitioning. Overall classification accuracy achieved by the segmentation method (88.33%) outperformed both per-pixel and ECHO by 1% and 6.66% respectively. Furthermore, the per-pixel method resulted in a large number of misclassification due to the 'salt-and-pepper' problem.

Comparisons may be improved through the recognition of structural differences between per-pixel and object-based methods. Radoux, Bogaert, Fasbender, & Defourny (2011) compared pixel and segment classification methods using mathematical simulations. Estimates for 1,000 and 10,000 sample units suggest that both bias and variance can be reduced in an object-based study (Radoux et al., 2011). They also recognise that pixel dimensions define the resolution and can be directly equated to comparable area (Radoux et al., 2011). Calculating accuracy for segmented outputs using an equation based on the assumption of equal area units may introduce uncertainty (Radoux & Bogaert, 2017). Therefore overall accuracy should be measured per the number of correctly classified segments (Congalton & Green, 2009).

Further benefits to an object-based imagery analysis is that uncertainties arising from positional discrepancies are likely to be reduced (Congalton & Green, 2009; Radoux et al., 2011). Geospatial locational errors may reduce the thematic maps quality due to an underestimation of the

root mean square error (Smith, Stehman, Wickham, & Yang, 2003). A point being measured in the real-world has higher probability of being located within a segment over being located within a pixel (Addink et al., 2007). Overall, the majority of comparative research suggests that an object-based image classification is superior to traditional per-pixel methods (Belgiu & Csillik, 2018; Jebur et al., 2014; Lu et al., 2010; Ma et al., 2017; Malahlela et al., 2014; Myint et al., 2011; Niphadkar et al., 2017; Radoux et al., 2011; Silveira et al., 2019; Wu et al., 2017; Zhang et al., 2018). Therefore the vegetation samples on DHI may be more accurately delineated using segments as the unit of classification.

2.3. Aerial imagery segmentation

Segmentation is the preliminary step for an object-based imagery classification which involves pre-determined shape thresholds to be set (Arbiol et al., 2006; Kawakubo et al., 2013). Meaningful classification of aerial imagery is achieved through segments which mimic the real-world features they are modelling (Burnett & Blaschke, 2003). The quality of the segmented output is correlated with the image object geometry and has been shown to influence the accuracy of the final classification outcome (Benz et al., 2004; Niphadkar et al., 2017; Zhang et al., 2012). Ma et al. (2017) conducted a systematic literature review of 254 object-based experiments within 173 scientific papers. They found that an optimal segmentation scale was significantly correlated ($p < 0.05$) with the spatial resolution of the imagery. Therefore high-resolution aerial imagery captured on DHI may provide a robust solution to object-based modelling. Whilst the potential for substandard segmentation is reduced due to level of detail captured using RPA imagery, we acknowledge that an unsuitable object area surface may skew the classification results. It is pertinent to pre-determine the segmentation method best suited for modelling vegetation at a species level to ensure confounding variables are mitigated.

The three main parameters for multi-resolution segmentation in eCognition are scale, colour, and shape (El-naggar, 2018; Fugara, Pradhan, & Mohamed, 2009; Johnson, Bragais, Endo, Magcale-Macandog, & Macandog, 2015; Möllera, Lymburner, & M.Volkc, 2007; Sertel & Alganci, 2015; Shao, Long, Liang, Chen, & Yuan, 2015). The scale parameter is related to the homogeny and heterogeneity of image features per the size of the object area calculated. A greater scale will result in a larger object size and vice versa. Scale tends to be the most ambiguous of the three parameters and therefore is the predominate focus of 'trial and error' classification assessments (El-naggar, 2018). Secondly, the colour parameter refers to the pixel values of the image and sets the threshold for clustering within the segmentation algorithm. Lastly, shape is a parameter that controls how tightly each discrete object will

be defined per the underlying raster image (Dekavalla & Argialas, 2018). This relates to the smoothness of the segment borders and their similarity to a bounding box and compactness as a function of the pixel clusters to perimeter edge (Zhang & Maxwell, 2006).

Attempts have been made to automate and semi-automate the process of segmentation and optimal parameter selection. Examples include region growing using spatial autocorrelation (Espindola, Câmara, Reis, Leonardo, & Monteiro, 2006; Zhang, Xiao, Song, & She, 2013), multi-band spectral angle (JianYang, Li, & He, 2014), iterative local variance measures (Drăguţ, Csillik, Eisank, & Tiede, 2014; Drăguţ, Tiede, & Levick, 2010), global heterogeneity and homogeneity measures (Johnson & Xie, 2011; Zhu, Cai, Liu, & Huang, 2016), Gaussian edge detection (Jing, Hu, Noland, & Li, 2012; Wang, Gong, & Biging, 2004), and multi-scale segmentation (Arbiol et al., 2006; Kim, Madden, & Warner, 2008; Salembier & Garrido, 1998; Tabb & Ahuja, 1997; Zhang, Jia, Li, Yuan, & Zhao, 2014). For each of the aforementioned examples, partitioning of imagery is dependent on the aerial image features and still requires some level of expertise to determine suitability of the segment calculation. There is no clear, definitive method that applies to every dataset (Singh, Singh, & Partridge, 2005). Determination of the segmentation algorithm applied will need to be based on the inherent qualities of the imagery being assessed.

Another identified issue with automated object area calculation is that the output may result in over- or under-segmentation (Jing et al., 2012; Kawakubo et al., 2013). Over-segmentation is the process of dividing a feature on the earth surface into an unreasonably large number of segments. Conversely, under-segmentation is when a real-world features is merged with proximal areas. For example a scale parameter of 200 and greater may generalise classes indistinctly whilst a scale of 20 may lose the inherent properties of the surface being modelled. Even if a final classification is numerically accurate, the rendered classes may not depict the underlying aerial imagery.

Multi-scale segmentation has been cited as the most common automated method for segmenting an image with the purpose of optimizing the segmentation scale (Ma et al., 2017). This method reduces over-segmentation of homogenous and under-segmentation of heterogeneous features associated with regional and local scales respectively (Johnson et al., 2015). However it requires remotely sensed capture to comprise both local and regional features (Arbiol et al., 2006; Kim et al., 2008; Zhang et al., 2014).

High-resolution RPA imagery classified to a species level is solely a localised study. The multi-resolution segmentation method in eCognition attempts to balance the diversity of image scale ranges by concentrating on the homogeny of local scales first before expanding to larger regions (Liu, Xu, Zhao, Yong, & Xin, 2015). Therefore, we propose a supervised semi-automated approach where user defined parameters in eCognition set thresholds for a "bottom-up" multi-resolution segmentation.

Research has shown that there may be a 'sweet-spot' for combinations of segment parameters and the accuracy they can achieve (Kavzoglu & Yildiz, 2014). Kavzoglu and Yildiz (2014) tested segment parameterisation for scales of 5, 10, 20, 35, 50, and 70; shape values of 0.1, 0.3, 0.5, 0.7, and 0.9; and compactness for 0.1, 0.3, 0.5, 0.7, and 0.9. Their aerial photo classification resulted in a maximum accuracy (94.13%) for scale and shape values of 20 and 0.2 respectively. However, combinations returned a pattern within the matrix where the top left boxes from an approximate diagonal line between a shape value of 0.7 and a scale value of 30 all returned a classification accuracy greater than 90%. The bottom right boxes showed an approximately linear reduction in accuracy until a scale of 70 and shape of 0.9 was reached. The scale of 70 and shape of 0.9 parameters showed the lowest accuracy value of 48.28%. Therefore, the method proposed in this paper to set the shape and compactness at a central point and adjusting the scale parameter may achieve a similar result to the time consuming alternative of adjusting all three independently.

A visual assessment of the object area's fit in relation to user-defined parametrisation has shown to provide a reliable form of segmentation suitability validation (Adam et al., 2016; Juniati & Arrofiqoh, 2017; Mathieu, Aryal, & Chong, 2007; Myint et al., 2011; Sertel & Alganci, 2015; Sibaruddin, Shafri, Pradhan, & Haron, 2018; Yan et al., 2007; Yu et al., 2006). For the purposes of assessing RPA capture on DHI we will compare determine the suitability of scale parameters from both a visual assessment and the statistical outcomes. This will allow conservation managers to discern the most appropriate scale of measurement and replicate the process for future vegetation monitoring.

2.4. Classification of segments using per-pixel based methods

Researchers have proposed a method to classify remotely sensed data by applying the per-pixel derived ML algorithm to segmented imagery (Liu & Yetik, 2010). Robinson et al. (2016) applied ML for detecting mesquite (*Prosopis* spp.) using segmented WorldView2 imagery captured along the Fortescue River Delta, north-west Pilbara, Australia. They found the method was effective for

measuring larger samples (overall accuracy of 93.2%), once smaller samples (e.g. seedlings and juveniles) had been removed (Robinson et al., 2016). However, the findings may be improved using a machine learning algorithm to classify the satellite imagery.

Studies have attempted to combine machine learning and ML methods for improved classification outcomes. Man, Dong, and Guo (2015) measured urban land use for the University of Houston using 0.74m spatial resolution Light Detection and Ranging data and 2.5m hyperspectral CASI satellite imagery (Man, Dong, & Guo, 2015). The results show that a mixed-model with both ML and SVM could outcompete a classification using solely SVM (94.7% to 87.6% overall accuracy respectively).

However, the ML classifier was developed for per-pixel based analysis and therefore is suitable for uni-modal datasets (Farrag, Megahed, & Darwish, 2019). Both the mesquite detection and urban land use assessment utilised high resolution satellite imagery. Remotely piloted aircraft imagery captured on Dirk Hartog Island may exhibit a different distribution of values due to capturing very high resolution data. When the spread of data is multi-model or collinearly distributed ML may not be a suitable classifier (Liua, Shia, & Zhang, 2011). A machine learning algorithm may be more equipped to process the high within class spectral variability found in sub-centimeter spatial resolution remotely sensed capture (Díaz-Varela, Iglesias, Castro, & Varelad, 2018).

2.5. Machine learning algorithms

In recent decades there has been a rise in machine learning algorithms within object-based imagery classification (Belgiu & Dra˘gut, 2016). Machine learning algorithms enable the ability for multi-variate predictions through the use of data driven decisions (Jordan & Mitchell, 2015; Magnussen, McRoberts, & Tomppo, 2009; Teluguntla et al., 2018; Thessen, 2016). Commonly cited machine learning techniques for imagery classification include neural networks (Sabanci, Kayabasi, & Toktas, 2017; Shao & Lunetta, 2012), decision tree algorithms (Waldner et al., 2016), Support Vector Machines (Alberto et al., 2016; Müllerová, Bartaloš, Brůna, Dvořák, & Vítková, 2017; Wu et al., 2017), Random Forest (Basukala, Oldenburg, Schellberg, Sultanov, & Dubovyk, 2017; Feng, Liu, & Gong, 2015), K-Nearest Neighbour (Tehrany, Pradhan, & Jebuv, 2013; Yu et al., 2006), phenological approaches (Czernecki, Nowosad, & Jabłońska, 2018; Klosterman et al., 2018; Nogueira et al., 2019), and mixed methods (Saboori, Torahi, & Bakhtayari, 2019; Salehi, Zhang, & Zhong, 2013). Here we

will assess the potential application of Convolutional Neural Network (CNN), SVM, and RF for analysing high-resolution aerial imagery captured on DHI.

A CNN is a deep learning algorithm which resembles the connectedness of data through end-to-end classification utilising supervised backpropagation for training (Wang et al., 2018). Ochoa and Guo (2019) applied CNN to measuring trees in high-resolution aerial imagery. The algorithm returned a classification accuracy of 98% (Ochoa & Guo, 2019). Whilst CNN was able to effectively classify vegetation, they found that the classification required large datasets and an extensive training time for the parameters. Object recognition may be achieved using CNN at the mercy of extensive parameterisation (Mostafa & Wang, 2019). The connectedness of a large parameter space in CNN may lead to overfitting (Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2014). To overcome this issue Pelt and Sethian (2018) factored a multi-scale approach by running simulations that incorporate dilated convolutions and dense connections. However, to achieve a multi-scale approach the imagery available requires both local and regional capture (Pelt & Sethian, 2018). This technique is not suitable when applied to high-resolution, localised aerial imagery.

Another non-parametric learning algorithm used to classify segmented imagery is the SVM. SVM is a supervised optimisation algorithm that builds a probabilistic model based upon an inputted training set (Burges, 1998; Vapnik, 1995). Studies have shown that the SVM method may handle multi-dimensional data and classify complex landscapes (Ghosh et al., 2014; Melgani & Bruzzone, 2004; Pal, 2005). There are some issues associated with the use of this machine learning algorithm. As the output quality directly relates to selection of appropriate training samples, human error may be unintentionally introduced into the model (Kumar et al., 2017; Mountrakis, Im, & Ogole, 2011). Furthermore, if classes are not linearly separable then additional kernelling is required (Mishra, Prasad, Kumar, Gupta, & Srivastava, 2017). This means that the methods required to achieve a suitable classification outcome may become unnecessarily complicated and incur higher time costs.

Numerous studies have compared SVM and the RF ensemble learning technique. Chen et al. (2018) used 2m resolution panchromatic and 8m resolution multi-spectral image captured by the Gaofen-1 satellite to compare SVM and RF. The results showed that RF computes the classification 46.61% faster than SVM. RF also equated a higher overall accuracy (88.16%) and Kappa co-efficient (0.827) than SVM (overall accuracy = 84.89% and Kappa co-efficient = 0.781) (Chen et al., 2018).

These findings are supported by Sesnie et al. (2008) who used Landsat TM bands and spectral indices to quantify forest types in northern Costa Rica. For both spectral reflectance and reflectance with indices, RF returned a higher accuracy than SVM (Sesnie et al., 2010). Adam, Mutanga, Odindi, and Abdel-Rahman (2014) studied the ability to derive thematic maps of coastal landscapes in KwaZulu-Natal, South Africa. RapidEye satellite imagery was clustered and then classified using RF and SVM techniques. Evaluation of the predicted outputs showed that both RF and SVM returned the same Kappa co-efficient of 0.92 (Adam et al., 2014). However, RF produced a slightly higher accuracy (93.07%) than the SVM algorithm (91.8%).

A few scientific papers have shown SVM to outperform RF when classifying regional areas using moderate to low-resolution satellite imagery (Noi & Kappas, 2017; Raczko & Zagajewski, 2017). However these studies were conducted on satellite imagery with a spatial resolution greater than 3m. The inherent complexity and heterogeneity of sub-centimeter resolution imagery may lend itself to overfitting and difficulties in parameterisation which is a known limitation of SVM (Breiman, 1984; Foody, 2004; Rodriguez-Galiano et al., 2012). The broader research outcomes shown suggest that RF may be more favourable than SVM for measuring the coastal vegetation found on DHI.

The RF algorithm is an ensemble classifier consisting of bootstrapped samples implemented in a Classification and Regression Tree (CART) analysis (Breiman, 2001; Cutler et al., 2007; Ho, 1998). This aggregate approach initially selects, at random, a subset of the training samples to develop individual classification rules. The ensembles then split the data into binary categorical predictions. Recursive partitioning of data into nodes occurs so that classes become increasingly homogenised as the most likely class is voted for. The extratrees split rule is a method of partitioning data based on the 'extremely randomised tree' principle which randomises the attribute and cut-point at each node (Geurts, Ernst, & Wehenkel, 2006). The strength of this method is based on the principle that the tree structure is independent of the output value. Forest structure will be completed once the CART has constructed a final decision on the majority class candidate.

An advantage of RF is that the algorithm uses adaptive bagging which may reduce bias through the ability to change the training set as it progresses (Breiman, 2001). Furthermore, convergence of the generalization error negates issues arising from overfitting (Breiman, 2001). Multiple studies have shown the effectiveness in computing aerial imagery classification using RF (Du et al., 2015; Jhonnerie

et al., 2015; Pantaleoni et al., 2009; Rodriguez-Galiano et al., 2012). Chapman, Bonn, Kunin and Cornell (2010) used RF to monitor vegetation in Peak District National Park, England using 5m spectral remotely sensed imagery. Their studied found that 97% of the 2,783 training points were accurately classified (Chapman et al., 2010). Studies classifying vegetation using RF have predominately analysed satellite imagery. Scientific literature applying an object-based paradigm with RF classification to sub-centimeter resolution RPA capture is scarce. As remotely sensed technology advances the capability to capture fine-details, so does the potential to derive geospatial information. Therefore, it is important that we determine suitable methods for extracting meaningful information from high-resolution aerial imagery for long-term monitoring of vegetation on DHI.

2.6. Ancillary variables

Remotely sensed imagery analysis may be improved via the exploitation of ancillary variables to multi-spectral capture. Multi-spectral in this instance refers to both the near-infrared band and the visual spectrum which comprises red, green, and blue wavelengths as individual bands (Olsen, 2007). These bands may be combined as a ratio or function to extract a greater understanding of the earth's surface (Babar, Ginkel, Klatt, Prasad, & Reynolds, 2006). The Green Leaf Area (GLA) algorithm exploits spectral reflectance information for delineating vegetation types and ground cover (Booth, Cox, Fifield, Phillips, & Williamson, 2005; Louhaichi, Borman, & Johnson, 2001). GLA values for an aerial image are calculated by the ratio of green, red, and blue wavelengths with additional weighting given to the green band (Macfarlane & Ogden, 2012).

Vegetation classification studies have traditionally leveraged the near-infrared band to delineate vegetation from other class types (Akar & Güngör, 2015; Fern, Foxley, Bruno, & Morrison, 2018; Huang, Liu, Li, Yan, & Ou, 2018; Jin, Liu, Chen, & Liang, 2018; Redowan, Akter, & Islam, 2014). However, not all RPA equipment has the ability to capture near-infrared wavelengths. Furthermore, whilst the near-infrared is beneficial for differentiating vegetation from other land cover types, it may not be as effective in separating plant species per their unique chlorophyll pigment reflectance properties. There is a research gap for studies testing the ability of GLA to classify high-resolution remotely sensed data compared to near-infrared indices. The GLA algorithm may provide scientists a suitable alternative for delineating a species level classification of RPA imagery.

In recent decades there have been a number of studies showing the effectiveness of GLA to classify remotely sensed imagery (Booth et al., 2005; Chianucci et al., 2016; Louhaichi et al., 2001; Macfarlane & Ogden, 2012). Chianucci et al. (2016) applied the GLA colour model to estimating Italian beech forest canopies using true colour RPA imagery with a 7.5cm spatial resolution. They found that the relationship between the RPA GLA planophile leaf angle distributions against a fisheye control image to validate the algorithm returned a coefficient of determination (R^2) of 0.7 (Chianucci et al., 2016). This suggests that GLA may provide a sound metric for estimating vegetation. The strength of the GLA algorithm in classifying vegetation per nadir pole camera images has been supported for estimates of Jarrah forest foliage cover ($R^2 = 0.99$) (Macfarlane & Ogden, 2012) and differentiating bare ground from plant communities to determine vegetation health across Colorado and Wyoming (Booth et al., 2005). However, there have not been any applications of the GLA algorithm in determining a species level classification of vegetation using high-resolution RPA imagery. The theory of using the GLA calculation as a substitute for near-infrared indices may be supported through further studies utilising the algorithm.

Height measurements may also improve accuracy outcomes when quantifying RPA imagery to a species level (Gleason, Shihavuddin, Gracias, Schultz, & Gintert, 2015). Studies have shown the usefulness of spot height data to discriminate between ground surfaces and other classes (Kluckner, Mauthner, Roth, & Bischof, 2009; Salehi, Zhang, Zhong, & Dey, 2012). Räsänen, Kuitunen, Tomppo, & Lensu (2014) tested whether a canopy height model was an important variable for determining boreal forest habitat types. Their method overlaid an airborne laser scanning canopy height model onto World-View 2 satellite multi-spectral imagery for an object-based classification using RF. Variable importance outcomes showed that the canopy height and canopy height standard deviation were highly ranked, especially when combined with texture measures (Räsänen, Kuitunen, Tomppo, & Lensu, 2014). However, there have been limited research classifying high-resolution imagery with height data where species is the unit of study. The height of a plant is also a characteristic field botanists employ when identifying vegetation to a species level (Barrett, 2016; Falster & Westoby, 2003). It is a logical step to apply tangible identification techniques to remotely sensed species classification.

Object-based methods allow for opportunities to utilise geospatial attributes when defining real-world entities. These may include the texture of an aerial image and the inherent geometric properties of imagery segments (Chen et al., 2018; Niphadkar et al., 2017). A common method to

determine the texture of an aerial image is by calculating a Grey-Level Co-occurrence Matrix (GLCM) (Davis, 2018; Wang et al., 2004). The GLCM recognises the X,Y spatial domain of a raster dataset and assigns an intensity value to each cell within the array (Coburn & Roberts, 2004; Haralick, Shanmugam, & Dinstein, 1973). It is based on the theory that texture and tone are inextricably linked (Haralick et al., 1973). Furthermore, the pixel distance and direction within an image kernel may hold additional information to extract spatial feature characteristics (Haralick et al., 1973). Incorporating texture into an object-based experimental design has shown to improve the classification results (Champion, Germain, Costa, Alborini, & Dubois-Fernandez, 2013; Feng et al., 2015; Kim, Warner, Madden, & Atkinson, 2011; Liu, Zhou, Zhou, Shao, & Yang, 2013; Wang et al., 2004).

Harlick et al. (1973) initially proposed 14 metrics for measuring texture using kernel based distance and direction to derive a classified land use map. Feng, Liu and Gong (2015) calculated six texture layers as ancillary data to the spectral reflectance values for classification of moderate resolution (0.07m) RPA capture. They found the inclusion of homogeneity, standard deviation, dissimilarity, entropy, mean, and angular second moment significantly improved the accuracy of classifying vegetation (Feng et al., 2015). The variable range for texture was not limited prior to classification. Incorporating a myriad of parameters for measuring texture may not always be feasible as texture calculations are computationally intensive (Shahbahrami, Pham, & Bertels, 2012). Temporal limitations within a project necessitates an apt selection of texture measures to mitigate the additional time taken to compute potentially redundant variables.

Kim et al. (2011) measured the spatial distribution of grey levels to determine salt marsh dieback using the Leica ADS40 RPA. Landscape capture measured features at a 0.3m pixel size. The method used a directional invariant kernel to calculate angular second moment, contrast, correlation, dissimilarity, entropy, homogeneity, mean, and variance. They found that only mean provided a dissimilarity between the features. Incorporation of the mean GLCM calculation improved the classification accuracy by 3-12% when compared with spectral bands alone. Mean may provide a reliable texture surface to discriminate plant features on DHI.

The use of mean calculations may be complemented by additional textures to ensure an exhaustive extraction of information from a remotely sensed image. Clausi (2002) used synthetic aperture radar to test the ability of GLCM statistics to classify sea-ice. The study showed contrast,

correlation, and entropy provided a comprehensive texture range and minimal variable set when compared with the use of one or all of the following: dissimilarity, uniformity, maximum probability, inverse difference, inverse difference normalized, inverse difference moment normalized, and inverse difference motion (Clausi, 2002). Deriving mean, correlation, contrast, homogeneity, and entropy texture features along with the spectral reflectance for classifying aerial imagery may improve classification accuracy.

Geometric spatial information may provide insights into dissimilarities between segment classes and improve the classification accuracy. Geometry or shape in an object-based image analysis is defined by the edge of the segment (Lillesand et al., 2015). Jiao and Liu (2012) propose that shape metric signatures may function similarly to spectral reflectance signatures when classifying land use class segments. Studies showing the pertinence of shape in delineating real-world features include classifying urban vegetation communities using Ikonos satellite imagery (Mathieu et al., 2007) and mixed land use segments from SPOT-5 imagery (Jiao & Liu, 2012; Memarian, Balasundram, & Khosla, 2013).

The incorporation of GLA, height, texture, and shape to complement spectral reflectance values for RPA classification may improve the accuracy of the object-based modelling. Therefore, these measures have been incorporated into the methods for deriving a species level classification on DHI.

3. Materials and methods

This chapter outlines the study area and data collection methods using a remotely piloted aircraft. The full framework for achieving the aim of very high-resolution species mapping from multi-spectral imagery is also provided.

3.1. Study area

Dirk Hartog Island (DHI) is located approximately 850km north of Perth and forms the western barrier of Shark Bay, Western Australia (Figure 3.1). The island geometry is elongated measuring approximately 80 km from the northernmost to southernmost point, 12.5 km greatest south to west width, and comprising 630km^2 of land. Two exclusion plots measuring 40 x 40 m were constructed in May 2018 to protect an area of vegetation from ground-dwelling fauna for the purpose of testing their impacts in comparison to external vegetation (Figures 3.2-3.3). This ensured the hare-wallabies would not interact with the experimental control (exclusion plot vegetation) and influence the results. Sites for the exclusion plots were located in the central and southern sections of the island denoted as plot x01 and x06 respectively. These sites were within areas of representative local vegetation. The materials used to construct the exclusion plots were star pickets and chicken wire standing at approximately 1 m high and extending 0.5 m outwards to ensure the perimeter is secure from ground-dwelling fauna. Assessments were conducted during autumn between the 29th of April, 2019 and 1st of May, 2019.

3.2. Geology, climate and vegetation

The island is topographically low-lying with an approximate maximal elevation of 180m above sea level (Whitlock, 1921). Geological assessments show the land comprises of predominately sandy dune systems with interdunal depressions and Tamala aeolian limestone outcrops (Guern & Davaud, 2005). Wind transported sand contains carbonate grains resulting from biogenic materials (shells, red algae, sea urchin spines, and foraminers) (Guern & Davaud, 2005). The climate is semi-arid with a strong southerly wind (Bowder, 1990; Harvey, Johnson, & Harvey, 2018). The closest weather station is Denham, WA estimating the average 1989-2018 rainfall to be 216.7mm per year (Bureau of Meteorology, 2019). During this period the mean minimum and maximum annual temperature was recorded as 17.8°C and 26.8°C respectively (Bureau of Meteorology, 2019). These conditions support five vegetation communities including tall open-heath, low closed-heath, low open-heath, hummock grassland, and low open shrublands (Maryan, 1996).

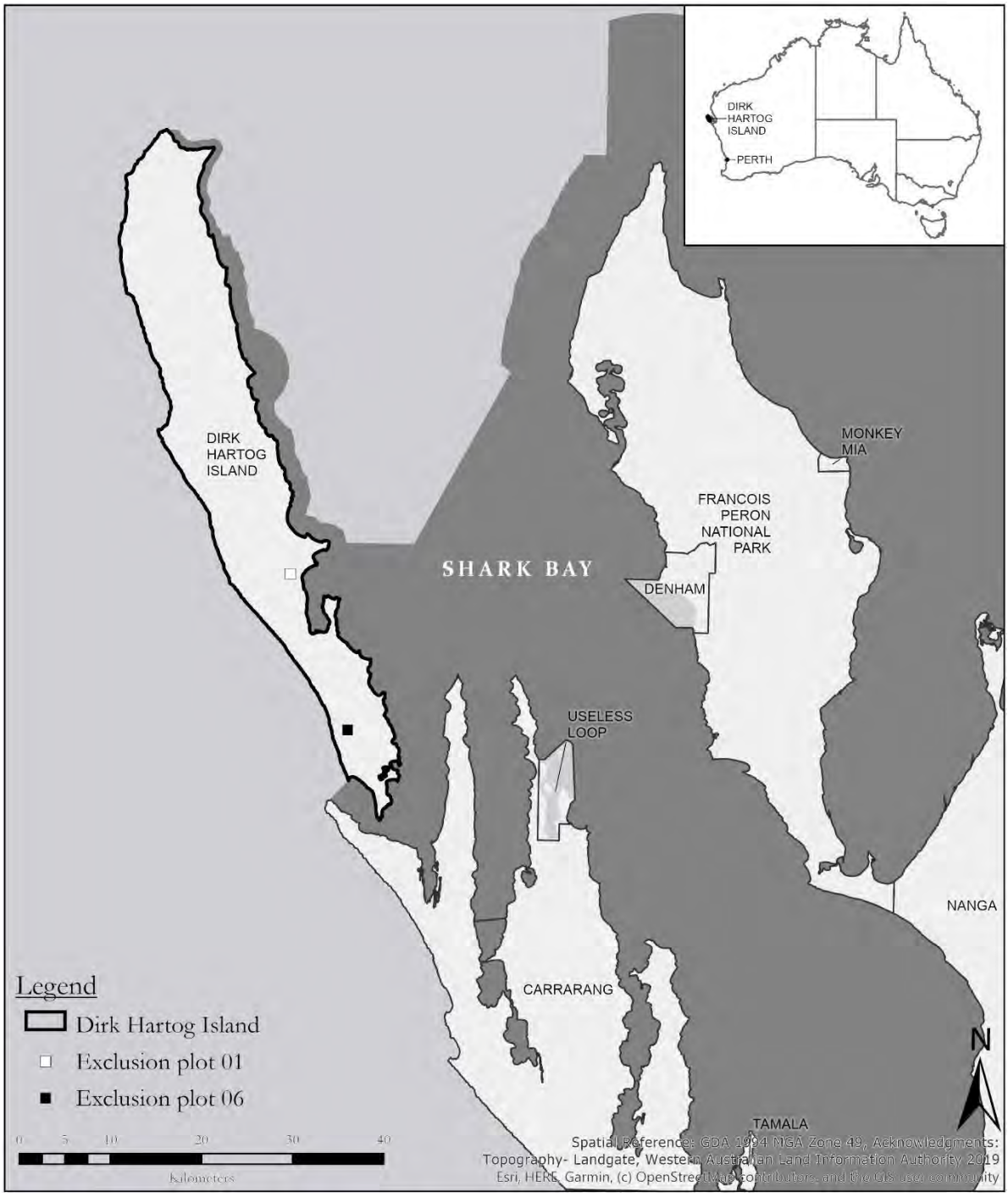

Figure 3.1. Map showing the location of Dirk Hartog Island and exclusion plot sampling sites. White square = exclusion plot 01, and black square = exclusion plot 06.

Figure 3.2. Northernmost exclusion plot site (ID=x01) map with remotely piloted aircraft capture. Red polygon = exclusion plot.

Figure 3.3. Southernmost exclusion plot site (ID=x06) map with remotely piloted aircraft capture. Red polygon = exclusion plot.

3.3. Research workflow

The framework for processing the Remotely Piloted Aircraft (RPA) imagery for detecting plant species is illustrated in Figure 3.4. This conceptual diagram shows the steps taken to derive a species level classification from the initial sampling of vegetation and RPA capture to producing a final classified output.

3.4. Field data capture

3.4.1. Geospatially locate and identify species

Flora species sampling was conducted between the 29th of April, 2019 and the 1st of May, 2019. Field researchers sampled within the exclusion plots and proximal areas. Plants were sampled from each study site on an incidental basis with the aim to capture at least 10 individuals per species, thus ensuring a viable sample size. Photographs were taken for each individual and the associative photo numbers were recorded. Approximate maximum height was measured using a graduated touch pole.

Attribute data for each sample (unique identifier, photo number, species, height (cm), and additional notes) were recorded using tangible field survey sheets. Samples were geospatially located using the Geocentric Datum of Australia (GDA) 1994, Map Grid of Australia (MGA) zone 49 co-ordinate system (Figure 4.3). Geospatially located sample centroids were estimated using a Samsung Android tablet. The unique identifier documented on the field survey sheet was also associatively recorded against the geospatially located sample. Field survey sheets were digitized into a Comma-Separated Value (CSV) file using Microsoft Excel version 15.0.5179.1000 ex-situ (Corporation, 2013). Additional records (20 for each exclusion plot site) were created in the CSV for the purpose of assigning a bare ground, seedlings, and juveniles class to samples at a later stage of the study. Each 'ground' record was assigned a unique identifier and given a height (cm) value of 0.

Identification of the vegetation to a species level was achieved using field guides provided from previous vegetation surveys. The field guides detailed likely species found within the selected sites and their biological characteristics. Diagnostic photographs of unknown individuals were sent to two individual botanists specializing in Western Australian taxa. A double blind approach was implemented to reduce bias and improve the reliability of their correct identification.

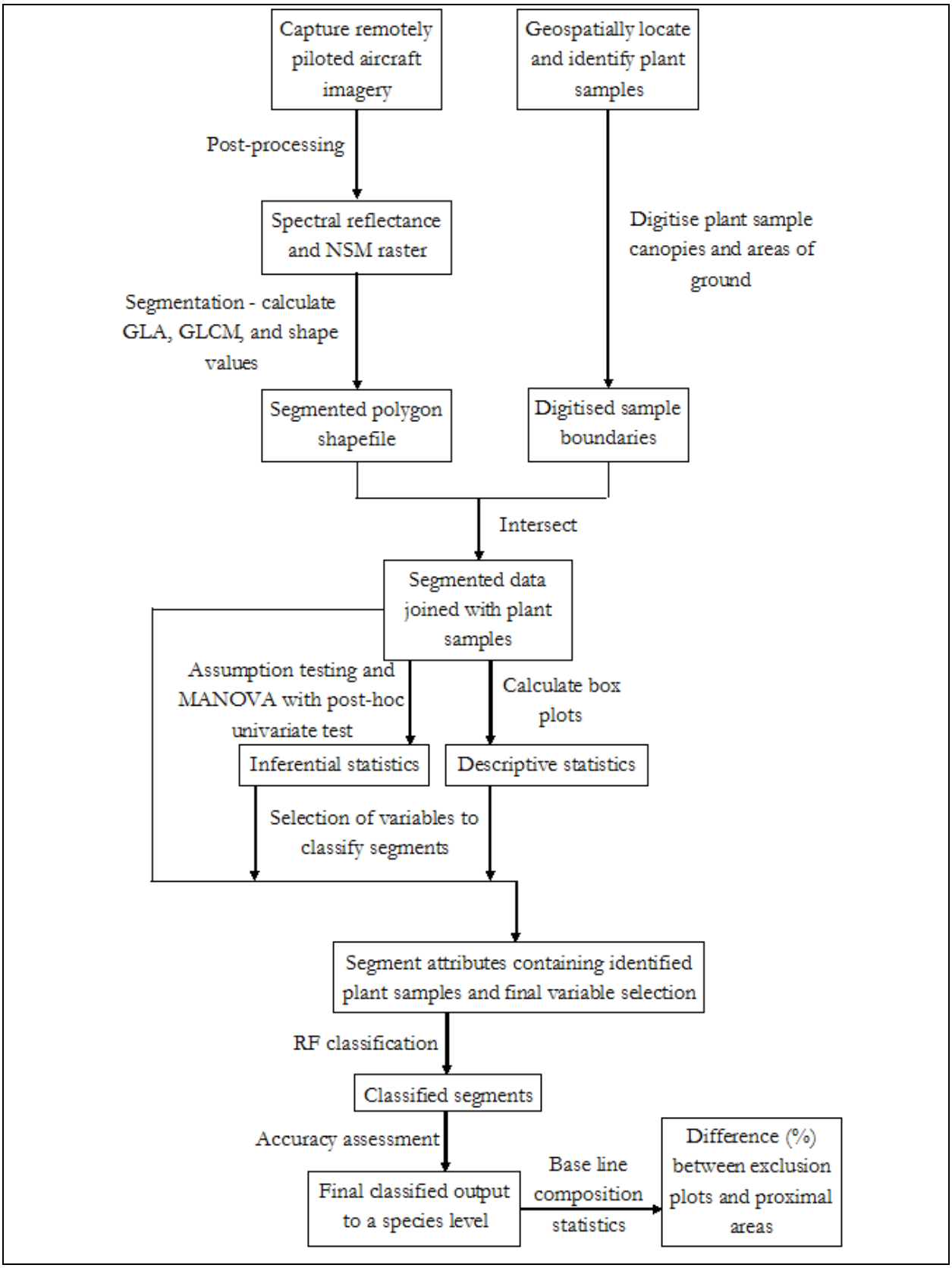

Figure 3.4. Conceptual diagram showing the workflow for classifying field data and measuring ecosystem engineered impacts.

3.4.2. Capture remotely piloted aircraft imagery

Images were captured using a DJI Phantom RPA on the 16th of September, 2018. Litchi version 4.14.0-g was used to determine a pre-recorded flight plan over the exclusion plot sites and to control the drone in the field (VC Technology, 2019). The flight plan sets the altitude of capture as the initial distance from the ground. For plot x01 and x06 the flying altitude was 11.9m and 25m respectively. Small surface undulations may impact the distance and provide slightly varying spatial resolution values. Due to the relatively flat surface found across both exclusion plot sites, these variations were deemed to be minimal and therefore would not affect the overall data quality. Ground resolution was captured as 5.49mm and 6.06mm per pixel across exclusion plot 01 and 06 respectively. Multi-spectral imagery was captured with an 8-bit radiometric resolution and centimeter positioning precision. Spectral reflectance was recorded for red, green, and blue wavelength bands. A Normalised Surface Model (NSM) was calculated during the post-processing phase of the study per height measurements captured by the drone. NSM values were assigned to each individual pixel in the form of a radiometric quantisation level (0-255) for x01 and an absolute height value for x06.

The RPA stored the capture as individual tiles using the World Geodetic System 1984 datum. Tiles were post-processed in Photoscan-pro version 1.4.2 (Agisoft, 2018). This allowed for a single mosaic image to be created for each site. The post-processed images were re-projected into the GDA 1994, MGA zone 49 spatial co-ordinate system.

3.5. Digitised sample boundaries

Geospatially located DHI vegetation sample points were overlaid onto the remotely piloted aircraft imagery in ArcGIS Pro version 2.4.0 (ESRI, 2019). Any species with less than 10 samples were removed from further modelling. Plant boundaries were digitised into polygons based on the approximated extent (without exceeding the boundary) shown in the aerial imagery. The plant boundary polygon was projected using the GDA 1994, MGA zone 49 spatial co-ordinate system. A total of 20 areas of ground, seedlings, and juveniles were digitised for each exclusion plot site and categorized as 'ground' within the plant boundary shapefile. A spatial join was implemented to associate the unique identifier of each geospatially located sample point captured in the field with a plant boundary polygon. 'Ground' polygons were manually assigned the unique identifier recorded in the comma-separated value file attributes previously mentioned.

3.6. Segmentation and variable extraction

Remotely piloted aircraft imagery was segmented using eCognition version 9.5.1 (Trimble, 2019). The extent of the segmentation was defined using a relative bounding box encapsulating the digitised plant boundaries. Shape and compactness parameters were set at 0.5 whilst variations of scale were assessed. Segment scale parameters were set at 20, 50, and 100, 200, and 400. A visual assessment of the parameterisation suitability was applied, in line with other studies segmenting aerial imagery (Adam et al., 2016; Juniati & Arrofiqoh, 2017; Mathieu et al., 2007; Myint et al., 2011; Sertel & Alganci, 2015; Sibaruddin et al., 2018; Yan et al., 2007; Yu et al., 2006). The scale value which best represented the underlying aerial imagery was selected for further statistical testing.

Spectral reflectance was derived from the aerial imagery and the mean red, green, and blue band values were calculated for each segment (Table 3.1). Reflectance was then used to derive the Green Leaf Algorithm (GLA) which is calculated by:

$$GLA = \frac{(2G - R - B)}{(2G + R + B)} \quad (3.1)$$

Where R = red wavelength value, G = green wavelength value, and B =blue wavelength value. This algorithm was applied to the RGB imagery and the result filtered using an 11x11 median filter to each pixel. The purpose of the filter was to smooth the GLA image to aid the segmentation of individual plant crowns. The object-based process calculates the maximum, median, and mean GLA value for each segment per the internal pixel values. Possible outliers were mitigated by calculating the 90th GLA percentile value per segment.

Class height was measured using mean, median, minimum, and maximum NSM values per each segment. Possible outliers were mitigated by calculating the 90th NSM percentile value per segment.

eCognition segmentation processing was used to compute the texture and shape metrics. Mean, correlation, contrast, homogeneity, and entropy grey-level co-occurrence matrices were computed in all directions as ancillary variables to the spectral reflectance, GLA, and NSM values (Haralick et al., 1973). Ancillary shape values calculated were the ratio of length and width, area per the number of pixels in each segment, roundness, and compactness for each segment.

Table 3.1. Variables derived per segment for classifying aerial imagery on Dirk Hartog Island.

Type	Variable	Field name
Spectral reflectance	Mean red band	Mean_red
	Mean green band	Mean_green
	Mean blue band	Mean_blue
NSM	Mean NSM	Mean_nsm
	Median NSM	Med_nsm
	Minimum NSM	Min_nsm
	Maximum NSM	Max_nsm
	NSM 90th percentile value	quan90nsm
GLA	Median GLA	Med_gla
	Mean GLA	Mean_gla
	Maximum GLA	Max_gla
	GLA 90th percentile value	quan90gla
Texture	Mean texture feature	GLCM_Mean_
	Correlation texture feature	GLCM_Corre
	Entropy texture feature	GLCM_Entro
	Contrast texture feature	GLCM_Contr
	Homogeneity texture feature	GLCM_Homog
Shape	Roundness	Roundness
	Compactness	Compactnes
	Length / width	LengthWidt
	Area per pixel	Area

In total there were 21 initial variables derived to delineate species within the exclusion plot sites. The final segmented shapefile consisted of segments in the form of contiguous polygons and the supporting aspatial data.

3.7. Data processing

Plant boundary polygons, CSV files containing field survey attributes, and the segment polygons for x01 and x06 were imported into R studio version 1.2.1335 implementing the R version 3.6.1 'Action of the Toes' coding language (R Studio, 2019; The R Foundation, 2019). A left join was applied from the CSV files to plant boundaries per the unique identifier captured in the field. Data was filtered into x01 samples and x06 samples. The library packages dplyr, stringr, sf, tidyr, raster were installed to process the inputted data (Hijmans, 2019; Pebesma, 2019; Wickham, 2019a, 2019c, 2019d). This process removed any additional note points not applicable to the study. The complete R code may be found in Appendix 9.1.

3.8. Statistical variable selection

3.8.1. Descriptive statistics

To summarise the dataset, the descriptive statistics calculated for each segment variable across both exclusions plot sites were range, mean, standard deviation, standard error, and skew. A visual comparison between species per variable was conducted via box and whisker plots, more commonly referred to as box plots. Box plots were created using the ggplot2 R library package (Wickham, 2019b).

3.8.2. Assumption testing

The underlying data contained multiple vectors of means and therefore a Multi-variate Analysis of Variance (MANOVA) was chosen. This statistical mechanism was proposed to test if the reduced variables, after assumption testing refinement, were significantly associated to species/ground types in order for the null hypothesis to be refuted. To determine if the data is suitable for a MANOVA test and to reduce the number of variables for machine learning classification, three assumptions needed to be met: 1) there were no correlations of greater than 0.6 between variables, 2) the data followed an approximately Gaussian distribution, and 3) the data relationships were approximately linear.

A correlation chart was derived to test assumptions 1, 2, and 3 per the Pearson method using the PerformanceAnalytics R library package (Nettleton, 2014; Peterson, 2019). Absolute Pearson correlation coefficient values show whether there is a correlation between two variables and the direction (±) of the correlation. A correlation greater 0.6, regardless of direction, was deemed to negate assumption 1. Where a correlation exists, the variable that is most dissimilar for species types per the

box plot comparison was chosen for analysis, under the assumption that it would provide the most unique information.

After assumption 1 testing was complete, skewness values of the remaining variable set were analysed to determine if assumption 2 had been met. If a variable returned a numerical skew value of >0.5 or <-0.5 a transformation was required. Variables with a skew value of <-0.5 were reflected prior to transformation. To apply a reflection, values were converted to negative numerals which created an inverse histogram. However, the logarithmic transformation requires that values are positive. Therefore, the maximum value plus 1 was added to each individual value within the variables where an inverse histogram had been calculated as seen in the following equation:

$$V_i = (V_1, V_2 \ldots V_n) + (V_m + 1) \quad (3.2)$$

Where V_i is the calculated value, V_n is the reflected value, and V_m is the maximum value for the variable.

A square root transformation was applied where the skew value was between 0.5 and 1 and a logarithmic transformation was applied to variables with a numerical skewness of greater than 1 (Webster & Oliver, 2001). The logarithmic transformation may not be applied to variables that contain zero values. For variables which require logarithmic transformation and contain the numeral zero, a constant of 1 was added to each individual value in the dataset.

The suffix '_TranF' was appended to the field title nomenclature for representation of transformed variables in this study. If both the square root and logarithmic transformation were unsuccessful for skew values of 0.5-1, the variable remained untransformed for further testing and the minor increase in the level of MANOVA uncertainty was accepted. If a logarithmic transformation could not stabilize a variable with a greater that 1 skew value, it was removed from further statistical testing. Data was stored in Microsoft excel spreadsheets using the xlsx R library package (Dragulescu & Arendt, 2018).

If a relationship did not meet the linearity assumption, the most dissimilar variable was removed from further statistical testing and classification per the box plot analysis.

Finally, to ensure there were no outliers confounding the outcome, an outlier plot was generated.

3.8.3. Multi-variate Analysis of Variance with post-hoc univariate test

A one-way MANOVA statistical test for significance was applied to the species/ground classes and the refined variables after assumption testing (Bray & Maxwell, 1985). The MANOVA was calculated per the Pillai's trace test (Pillai, 1955). If the MANOVA was found to be significant a post-hoc test was required. To determine which individual variables were significant for delineating vegetation on DHI, a univariate post-hoc test was applied through a Summarise Analysis of Variance model. Significance was tested at a $P < 0.05$ level throughout the analysis.

3.9. Random Forest classification

Random Forest (RF) was the machine learning algorithm chosen for high-resolution remotely piloted aircraft imagery classification. This is due to the algorithms ability to compute complex inputs and the inbuilt convergence of generalization error which mitigates overfitting issues (Breiman, 1984; Foody, 2004; Rodriguez-Galiano et al., 2012). Parameterisation of the model required an mtry value, minimum node size, and the number of folds. For comparison of dataset inputs the parameters for each run were set as mtry = 2, minimum node size = 1, and number of folds = 10. The number of folds value of 10 determines that 10% of the data would be portioned for the out of bag error estimate. The model was coded using the ranger and caret R library packages (Kuhn, 2019; Wright, 2019).

The machine learning ensemble classifier was run on the selected scale segmented x01 and x06 data for all variables, the selected raw variables per the MANOVA calculation, and the selected dataset with transformed variables per the MANOVA compare methodological outcomes. The RF algorithm was also run on the segmentation scales directly below and above the chosen segment scale parameter per all variables for both exclusion plot sites. Ten iterations per scenario was calculated and the out of bag error estimates quantified per mean Kappa and accuracy values were recorded for each. The approximate time to compute the scenario was also recorded.

The final iteration for the best suited accuracy outcome was exported as a shapefile with a populated predicted class field.

3.10. Exclusion plot and proximal area comparisons

Quadrat dimensions of 15 x 15 m were derived, both within the exclusion plot area and a proximal area, for each of the exclusion plot sites. The purpose of measuring quadrats is to compare vegetation cover per the predicted segment classes. Vegetation cover was used as a proxy for testing if the re-introduced fauna are acting as ecosystem engineers. The quadrat locations were determined by the segmentation bounding box in relation to the exclusion plot boundary. The cardinal direction with the maximum proximity testing area of segments was chosen as the location of the proximal quadrat area. A quadrat was calculated within the exclusion plot and the mirrored external area. Quadrats were measured 1 m from the exclusion plot fence line. The position along the fence line was chosen using a randomly generated longitudinal ordinate for the north/south fence line boundaries and latitudinal ordinate for the east/west fence line boundaries. The segmentation shapefile containing the predicted field for each site was intersected with the quadrat to calculates class cover (area per pixel and %) and compare outcomes.

4. Results

4.1. Remotely piloted aircraft imagery

A total of 284 and 314 individual photos were captured across exclusion plot x01 and x06 respectively. There was a visible increase in vegetation density and cover for exclusion plot site x06. In contrast, exclusion plot site x01 had a greater level of bare ground coverage per a visual analysis.

4.2. Vegetation samples

Incidental surveying of vegetation on Dirk Hartog Island (DHI) was found to be a proficient sampling method (Figure 4.1-2). Flora species sampling for x01 was conducted on the 29/04/2019, 30/04/2019, and 1/05/2019. Site x06 was sampled on the 30/04/2019. A total of 179 and 82 plant individuals were sampled across x01 and x06 respectively (Figure 4.3, Appendix 9.2).

The species found at x01 were *Acacia ligulata*, *Acanthocarpus preissii*, *Alyogyne cuneiformis*, *Atriplex vesicaria*, *Cenchrus ciliaris*, *Exocarpus aphyllus*, *Santalum acuminatum*, *Threlkeldia diffusa*, *Triodia plurinervata*, *Westringia dampieri*, and an unknown plant recorded as x01-008. Vegetation samples recorded as x01-008 were unable to be identified visually by trained botanists due to an absence of fruit or flowers. The plant structure and leaf morphology suggest that the samples were related. For the purposes of a species level grouping to classify aerial imagery the vegetation set x01-008 was deemed acceptable for inclusion in the object-based model.

S. acuminatum returned 1 individual specimen and *W. dampieri* had a total of 4 samples for exclusion plot site x01. *S. acuminatum* and *W. dampieri* were removed from further modelling due to falling beneath the 10 samples per species threshold.

There was a decrease in richness found at x01 with the four species identified being *Scaevola spinescens*, *A. ligulata*, *A. vesicaria*, and *T. plurinervata*. All species had a total sample count of >= 20.

Figure 4.1. Sampling effort map for the northernmost exclusion plot (x01) with remotely piloted aircraft capture. Red polygon = exclusion plot, green polygon = segment area, yellow points = individual sampled on the 29th of April 2019, blue points = individual sampled on the 30th of April 2019, and pink points = individual sampled on the 1st of May 2019.

Figure 4.2. Sampling effort map for the southernmost exclusion plot (x06) with remotely piloted aircraft capture. Red polygon = exclusion plot, green polygon = segment area, and blue points = individual sampled on the 30th of April 2019.

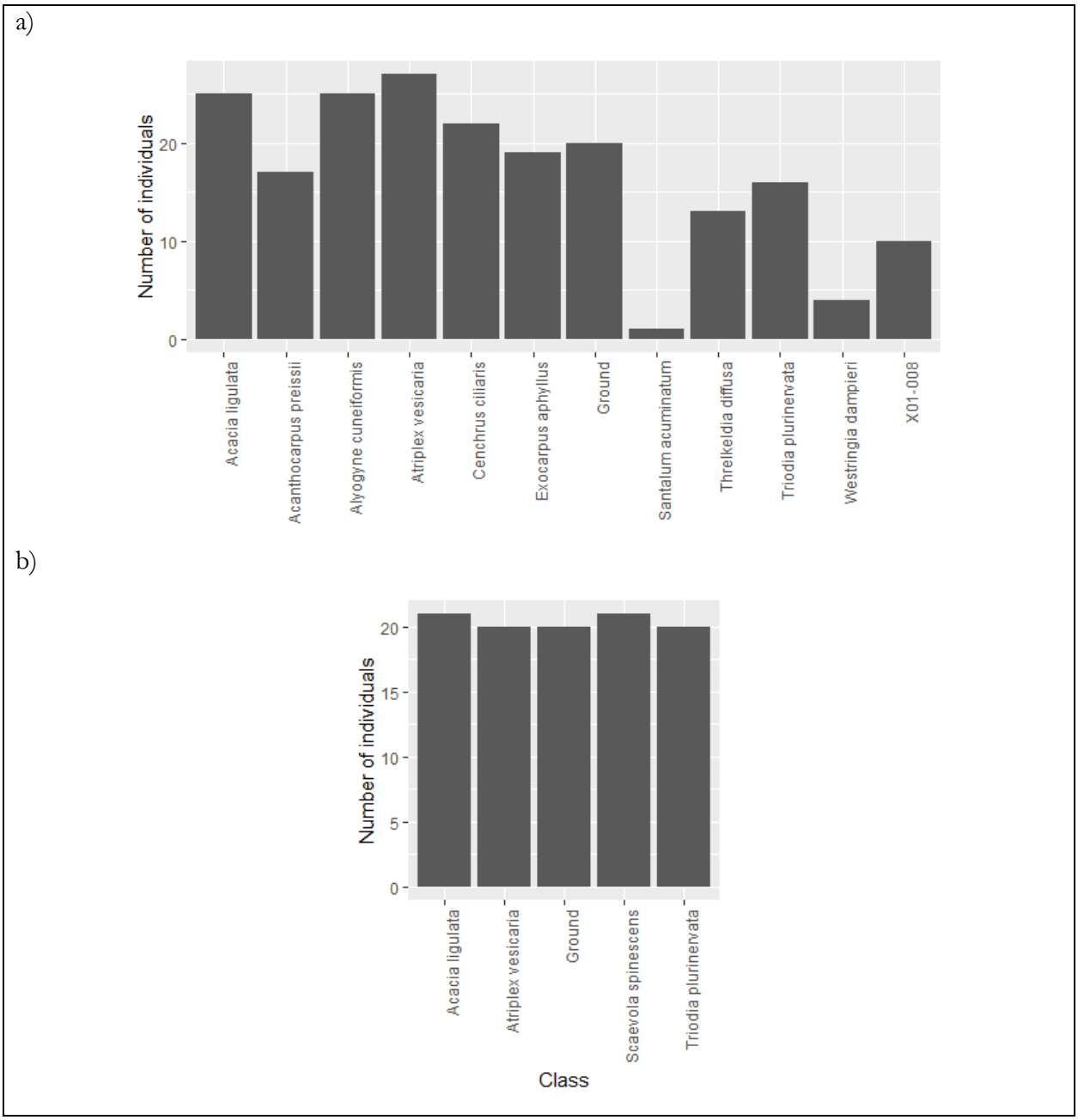

Figure 4.3. Distribution of samples captured from exclusion plot sites 1 and 6 where the x axis = species name/ground class and y axis = total number of individuals sampled. a) ex01 and b) ex06.

4.3. Segmentation output

The visual inspection of segmentation outputs showed that a scale of 50 best represented the underlying imagery for both exclusion plot sites. Segmentation for exclusion plot sites 1 had a total area of 4,215.1m^2 and exclusion plot 6 had a total area of 3,000.66m^2. There were 165,935 and 104,800 polygons delineated for x01 and x06 respectively.

Segment data intersected with known plant boundaries showed *A. ligulata* and *A. cuneiformis* presented the highest number of segments for x01 totaling 2,962 and 2,783 respectively (Table 4.1). *A. vesicaria* followed as the third most segments for x01 with a total of 1,233. *A. ligulata* segments for x06 also showed the highest count with 2,098 in total. *A. cuneiformis* was not found at x06. The subsequent segmentation total after *A. ligulata* was *S. spinescens* with 1,467 segments for x06. *A. vesicaria* had the second lowest segmentation count for x06 with 790 in total. *C. ciliaris* showed the lowest number of segments with 230 in total for x01. This species was not found at x06. *T. plurinervata* returned the lowest segmentation count for x06 with 524 polygons and the third least for x01 with 310 polygons.

Table 4.1. Segmentation counts per plant species/ground classes for x01 and x06.

Class	Total number of segments	
	x01	x06
Acacia ligulata	2,962	2,098
Acanthocarpus preissii	903	-
Alyogyne cuneiformis	2,783	-
Atriplex vesicaria	1,233	790
Cenchrus ciliaris	230	-
Exocarpus aphyllus	1,637	-
Ground	2,707	899
Scaevola spinescens	-	1,467
Threlkeldia diffusa	585	-
Triodia plurinervata	310	703
X01-008	308	-

4.4. Descriptive statistics variable comparison

Spectral reflectance were considered as the Mean_Red, Mean_Green, and Mean_Blue variables and shared similar class distributions for both x01 and x06 (Figures 4.4-5). *A. ligulata*, *A. cuneiformis*, and *E. aphyllus* spectral reflectance centers were positioned at relative intervals to each other and lower than the other plant species types for exclusion plot x01. The Mean_green variable for x01 species *A. ligulata*, *A. cuneiformis*, and *E. aphyllus* had a median value which was positioned slightly lower than other classes.

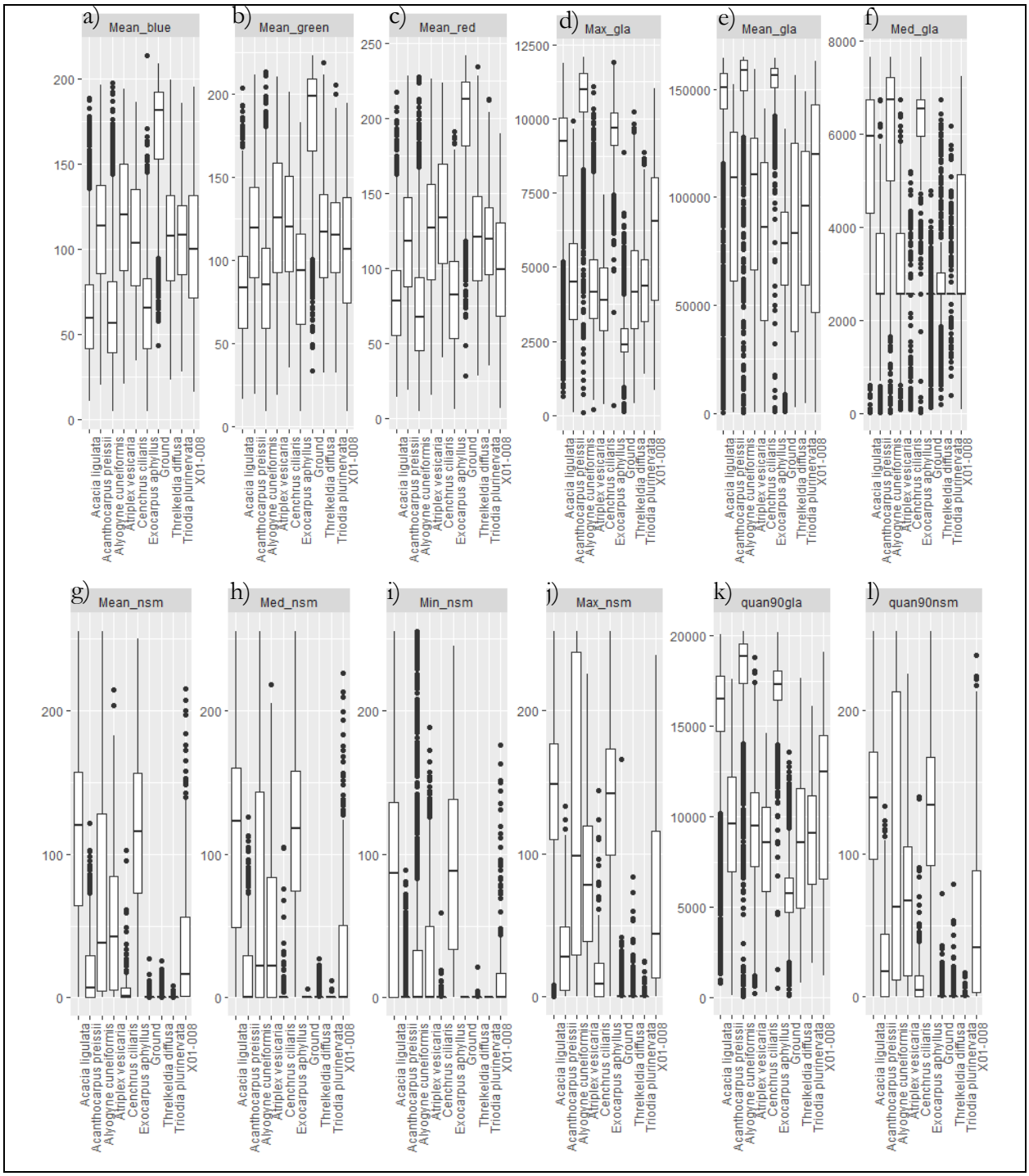

Figure 4.4. Box and whisker plot comparisons showing the spread of segment variables quantified per species and ground classes for x01. The whiskers encompass samples within 10% to 90% of each distribution with points representing values outside of this range. a) Mean_blue, b) Mean_green, c) Mean_red, d) Max_gla, e) Mean_gla, f) Med_gla, g)Mean_nsm, h) Med_nsm, i) Min_nsm, j) Max_nsm, k) quan90gla, and l) quan90nsm.

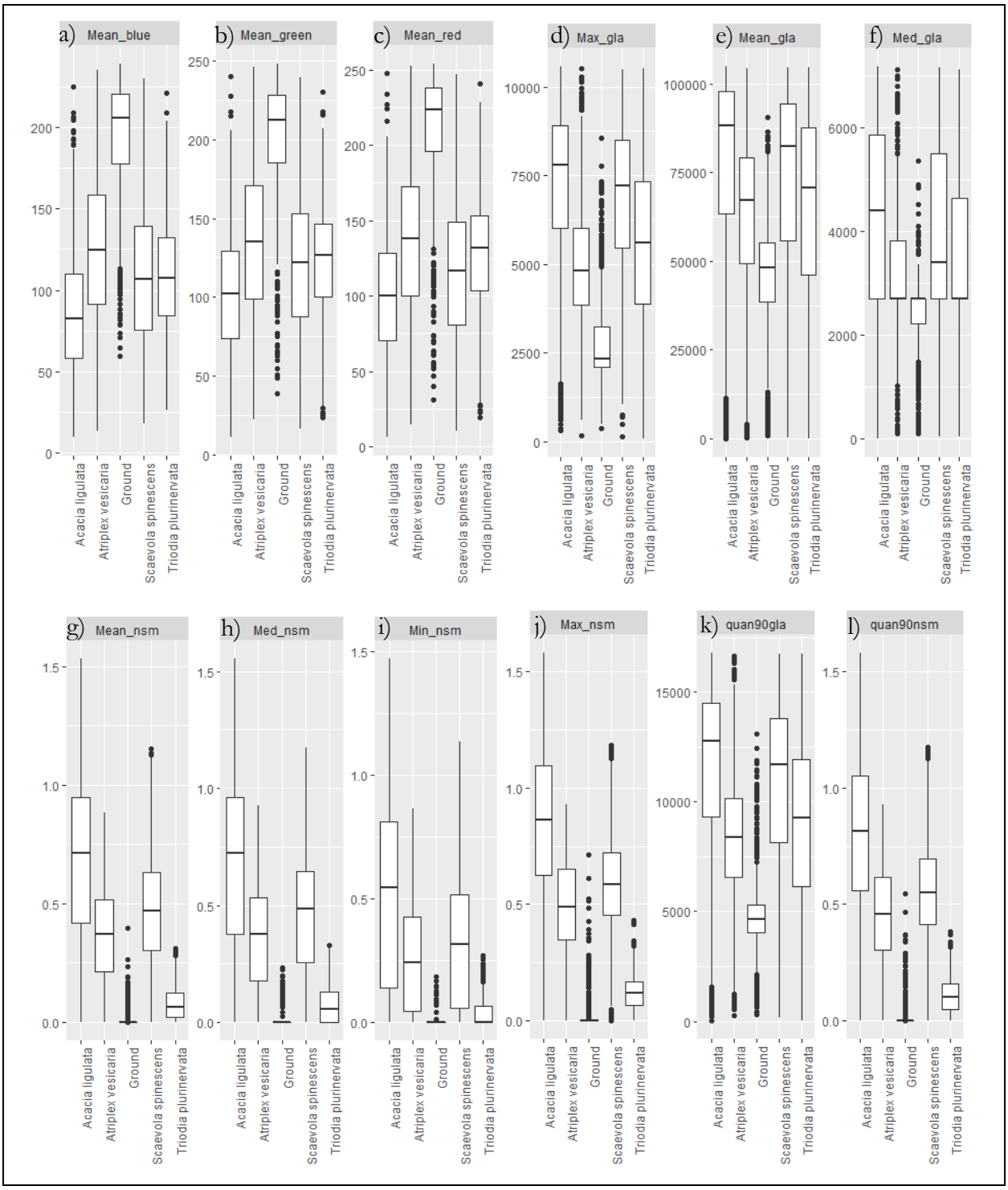

Figure 4.5. Box and whisker plot comparisons showing the spread of segment variables quantified per species and ground classes for x06. The whiskers encompass samples within 10% to 90% of each distribution with points representing values outside of this range. a) Mean_blue, b) Mean_green, c) Mean_red, d) Max_gla, e) Mean_gla, f) Med_gla, g)Mean_nsm, h) Med_nsm, i) Min_nsm, j) Max_nsm, k) quan90gla, and l) quan90nsm.

A. ligulata, *A. cuneiformis*, and *E. aphyllus* did share considerable interquartile range overlap with other classes. Alternatively, Mean_red and Mean_blue *A. ligulata*, *A. cuneiformis*, and *E. aphyllus* interquartile ranges were positioned slightly lower and shared minimal overlap with other classes for x01.

A. ligulata centers measured for x06 were slightly lower for each of the spectral reflectance variables. The third quartile of *A. ligulata* overlapped entirely with the second quartiles of other plant classes for all x06 spectral reflectance variables.

A. vesicaria has a slightly greater spread of interquartile range values and slightly higher center when compared with other plant species for both x01 and x06. The x01 spectral reflectance variables for *A. vesicaria*, *A. preissii*, *C. ciliaris*, *T. diffusa*, *T. plurinervata* and x01-008 showed an approximately Gaussian distribution with overlap between their interquartile ranges. Plant class x01-008 sampled in x01 showed a center positioned slightly lower than other mid-ranged plant species (*A. preissii*, *A. vesicaria*, *C. ciliaris*, *T. diffusa*, and *T. plurinervata*) for Mean_red. x01-008 shared a similar middle positioning interquartile range and center when compared to the other plant species for Mean_green and Mean_blue.

A normal distribution pattern across all spectral reflectance variables with an overlap between their interquartile ranges was found for the following x06 plant species: *S. spinescens*, *A. vesicaria*, and *T. plurinervata*. The latter two species shared an approximate center for Mean_blue and Mean_green. However, the center for x06 *S. spinescens* was slightly higher than *T. plurimervata* for Mean_red. The median value for *A. vesicaria* was positioned slightly higher than that of *S. spinescens* or *T. plurinervata* for all spectral reflectance variables.

Ground was the most dissimilar class with a notably higher inner quartile range (above 150) for all spectral reflectance variables across both exclusion plot sites. The spectral reflectance variable Mean_red followed closely by Mean_blue showed the most prominent dissimilarity for both exclusion plots.

It was noted that *A. ligulata* showed outliers at the upper tail for the spectral reflectance variables across both x01 and x06. *A. cuneiformis* followed the same spectral reflectance outlier pattern

for x01. These were most likely related to the presence of a possible skew. Further skew testing is required to test this assumption.

The Green Leaf Algorithm (GLA) calculations incorporated the variables Med_GLA, Mean_gla, Max_gla, and quan90gla. Each calculation appeared to contain numerous outliers for x01. There was a presence of outliers across each of the x06 GLA variables for all species excluding *S. spinescens*, and *T. plurinervata*. The index calculates the maximum value into the thousands and therefore creates a sparsity of data within the range. This sparsity may interpret data positioned along whiskers as outliers.

A. ligulata, *A. cuneiformis*, and *E. aphyllus* classes had higher interquartile ranges with no overlap when compared with the other classes for all x01 GLA variables except Med_gla. *A. ligulata*, *A. cuneiformis*, and *E. aphyllus* had higher centers for x01 Med_gla when compared with other classes. However, there was some x01 Med_gla overlap between the second quartiles of *A. ligulata* and *A. cuneiformis* and the third quartile of x01-008. Furthermore, there was no difference between the centers of *A. preissii*, *A. vesicaria*, *C. ciliaris*, *T. diffusa*, *T. plurinervata* and x01-008 for x01 Med_gla.

A. preissii, *A. vesicaria*, *C. ciliaris*, *T. diffusa*, *T. plurinervata* and x01-008 showed overlap between the interquartile ranges for all x01 GLA variables. The ground class was positioned relative to *A. preissii*, *A. vesicaria*, *C. ciliaris*, *T. diffusa*, *T. plurinervata* and x01-008 for all x01 GLA variables except for Max_gla. Max_gla which showed the ground interquartile range was lower than the other plant species with no overlap for x01.

There was some quartile range overlap between the plant classes for each GLA variable across x06. *A. ligulata* had the highest center for all x06 GLA variables followed by *S. spinescens*. The ground class was moderately positioned relative to other plant species for all x06 GLA variables excluding Max_gla and quan90gla. Max_gla and quan90gla for x06 showed the ground interquartile range was lower than the other plant species with no overlap.

The distribution of Max_gla and Mean_gla values appeared approximately Gaussian for all classes with the exception of *A. ligulata*, *A. cuneiformis*, and *E. aphyllus* for x01 and *A. ligulata* and *S. spinescens* for x06. Med_gla for both exclusion plots showed the least similarity. Max_gla followed by

Mean_gla were visually determined to portray the most dissimilarity between classes for both x01 and x06 GLA variables.

The Normalised Surface Model (NSM) variables included Med_nsm, Mean_nsm, Min_nsm, Max_nsm, and quan90nsm which had differences in values between x01 and x06. NSM variables showed similar overall class distributions for both x01 and x06. However, distribution of values shared similar characteristics between the exclusion plot sites. *A. ligulata* showed a normal distribution and the highest upper whisker across each NSM variable for x06. Similarly, *A. ligulata* had one of the highest upper whiskers for x01 along with *A. cuneiformis* and *E. aphyllus* across each of the NSM variables. The spread of *A. ligulata* and *E. aphyllus* was relatively normal across Med_nsm, Mean_nsm, Max_nsm, and quan90nsm for x01. However, *A. ligulata* showed a strong positive skew for Min_nsm in x01. *E. aphyllus* was approximately normally distributed for Min_nsm in x01. *A. cuneiformis* showed a strong positive skew in all NSM variables except Max_nsm and quan90nsm for x01.

Across exclusion plot site x06 *A. vesicaria* and *S. spinescens* followed an approximately parallel interquartile range for all NSM variables. However, for all x06 NSM variables *A. vesicaria* showed a slightly lower upper whisker compared to *S. spinescens*. *T. plurinervata* showed minimal interquartile range overlap when compared with *A. vesicaria*, *S. spinescens*, and ground across all x06 NSM variables except Min_nsm. The ground class, which also incorporates seedlings and juveniles, shows the best representation of values within the Med_nsm and Min_nsm variables for both exclusion plot sites. Based on the overall distribution of values and the representation of the ground class, Med_nsm may be most suitable for RPA imagery classification followed closely by Min_nsm for both x01 and x06.

Texture surfaces calculated were the GLCM_Mean_, GLCM_Corre, GLCM_Entro, GLCM_Contr, and GLCM_Homog variable set (Figure 4.6-7). GLCM_Mean_ shows similar Gaussian distributions across all classes for both exclusion plot sites.

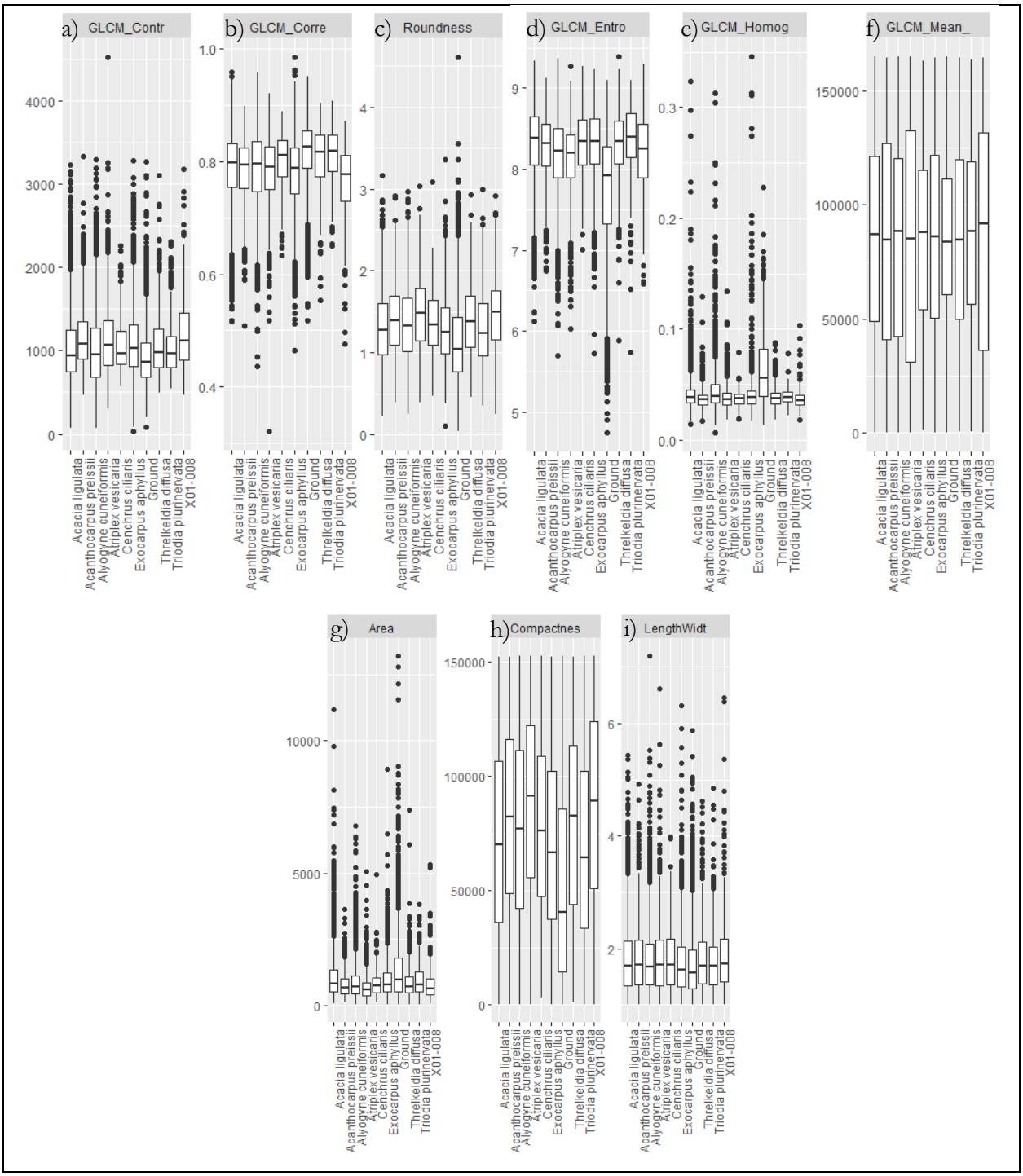

Figure 4.6. Box and whisker plot comparisons showing the spread of segment variables quantified per species and ground classes for x01. The whiskers encompass samples within 10% to 90% of each distribution with points representing values outside of this range. a) GLCM_Contr, b) GLCM_Corre, c) Roundness, d) GLCM_Entro, e) GLCM_Homog, f) GLCM_Mean_, g) Area, h) Compactnes, and i) LengthWidt.

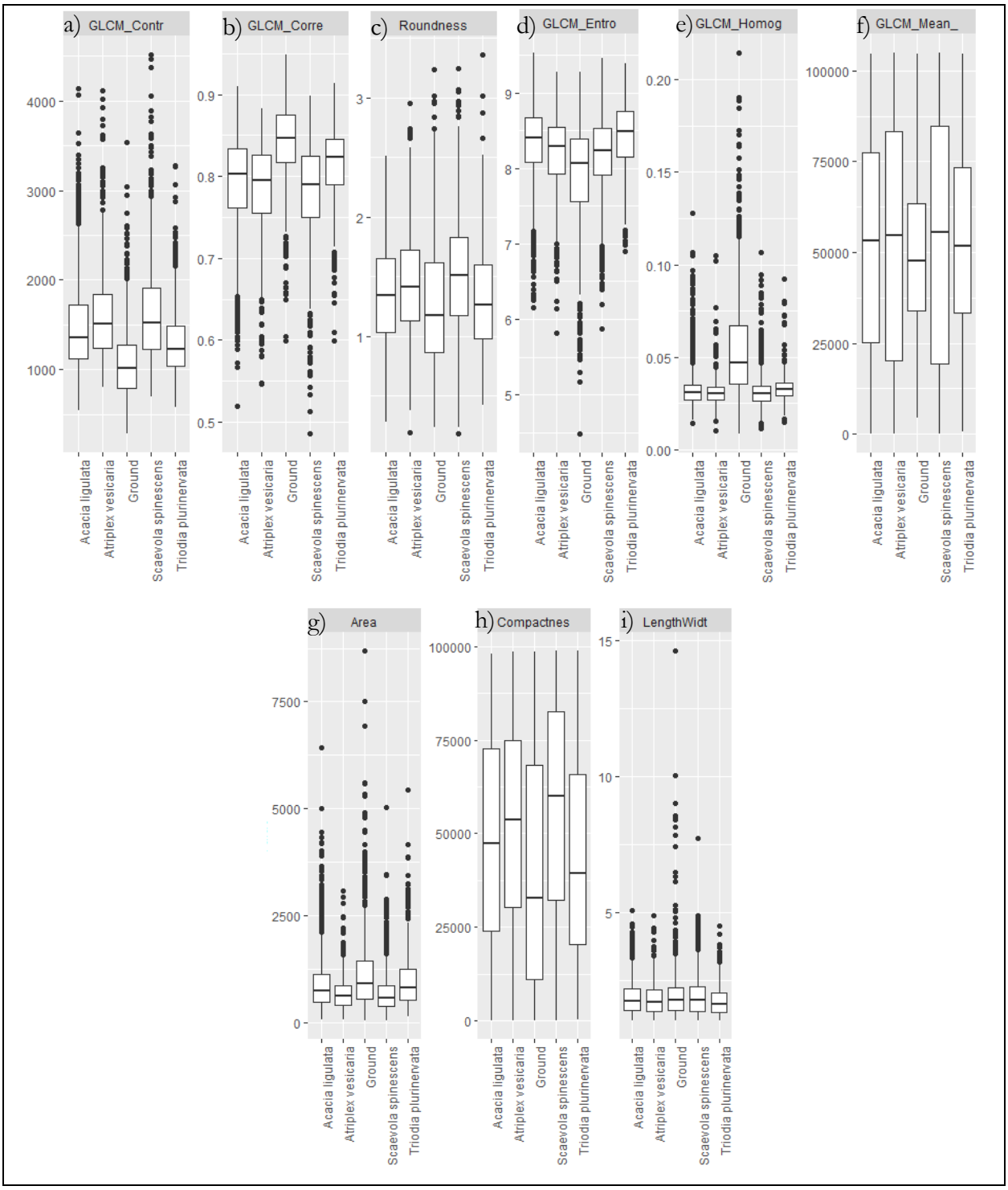

Figure 4.7. Box and whisker plot comparisons showing the spread of segment variables quantified per species and ground classes for x06. The whiskers encompass samples within 10% to 90% of each distribution with points representing values outside of this range. a) GLCM_Contr, b) GLCM_Corre, c) Roundness, d) GLCM_Entro, e) GLCM_Homog, f) GLCM_Mean_, g) Area, h) Compactnes, and i) LengthWidt.

GLCM_Homog showed a strong positive skew for both x01 and x06. There was no conspicuous GLCM_Homog difference between the interquartile range and spread between plant species for both x01 and x06. However, the ground class may be discerned from plant species for both exclusion plot sites per a non-overlapping GLCM_Homog interquartile range. GLCM_Homog may delineate ground to surrounding vegetation for both exclusion plot sites.

GLCM_Corre and GLCM_Contr showed no discernable difference between classes for x01. Plant species distributions were relatively similar across both GLCM_Corre and GLCM_Contr for x06. However, the median and interquartile range for the ground class were discrete from the plant species center values. GLCM_Corre and GLCM_Contr had slightly varying plant species centers for x06. GLCM_Contr showed a slightly greater dissimilarity between interquartile ranges when compared with GLCM_Corre for x06. GLCM_Contr and GLCM_Corre for x06 may assist in improving the classification accuracy for species delineation on Dirk Hartog Island.

Both exclusion plot sites showed a similar partitioning of plant species values for GLCM_Entro. GLCM_Entro showed an extended lower whisker for the ground class with minimal interquartile range overlap when compared with the plant species classes for x01. The subtle differences between ground and plant species found in GLCM_Entro for x01.

Auxiliary shape variables include Roundness, Compactnes, LengthWidt, and Area. Roundness showed relatively no differences between the centers and ranges of classes for both x01 and x06. There was a presence of Roundness outliers for all x01 and x06 classes with the exception of *A. ligulata* sampled for x06. Compactnes values were approximately normally distributed across plant species with a minor positive skew shown in the ground class for both exclusion plot sites. Compactnes showed minimal variation between class centers for both exclusion plots. Furthermore, there was a moderate decrease in the ground class center compared with plant species for both x01 and x06. *A. vesicaria* and X01-008 Compactnes box plots showed a slightly higher center when compared with other x01 classes. Compactnes values for x06 show *S. spinescens* as having the highest center.

Area box plot calculations for both exclusion plots sites show a strong positive skew across all classes. There did appear to be some Area variation in the interquartile range and upper whiskers between classes for both x01 and x06. However, the skew made it difficult to determine if Area could

differentiate classes within the Random Forest algorithm. The Area variable for both exclusion plot sites may improve classification accuracy is data is transformed and becomes normally distributed.

LengthWidt showed a strong positive skew for all classes in x01 and x06. It was determined that LengthWidt showed no difference in interquartile range and center between classes for x01 or x06.

From the geometry variables available, Compactnes may complement other non-shape variables when classifying segments derived from aerial imagery capture across x01 and x06. Area for both exclusion plots may provide useful information for delineating vegetation if transformed prior to Multi-variate Analysis of Variance testing.

4.5. Inferential statistics variable comparison

4.5.1. Assumption 1 testing - correlations

The correlation matrix for exclusion plot site 1 shows a high correlation between spectral reflectance variables (Figure 4.8-9). Mean_blue and Mean_green showed a correlation of 0.97, Mean_blue and Mean_red showed a correlation of 0.98, and finally Mean_red and Mean_green showed a correlation of 0.99. Therefore, retaining all three variables in the Multi-variate Analysis of Variance (MANOVA) test would not satisfy assumption 1, that correlations are not greater than 0.6. The box plot assessment suggested that the dissimilarity between Mean_red classes was greater than that of Mean_blue or Mean_green for exclusion plot site x01. The Mean_red variable was selected for further exclusion plot site x01 testing.

Exclusion plot site x06 Mean_red, Mean_green, and Mean_blue showed an analogous correlation pattern to the spectral reflectance found across x01. Mean_blue and Mean_green showed a correlation of 0.98, Mean_blue and Mean_red showed a correlation of 0.97, and Mean_red and Mean_green showed a correlation of 0.99. Mean_red showed the greatest dissimilarity between classes via a box plot analysis. To mitigate the correlations found in the x06 matrix and satisfy assumption 1, Mean_red was selected from the spectral reflectance variables for further testing.

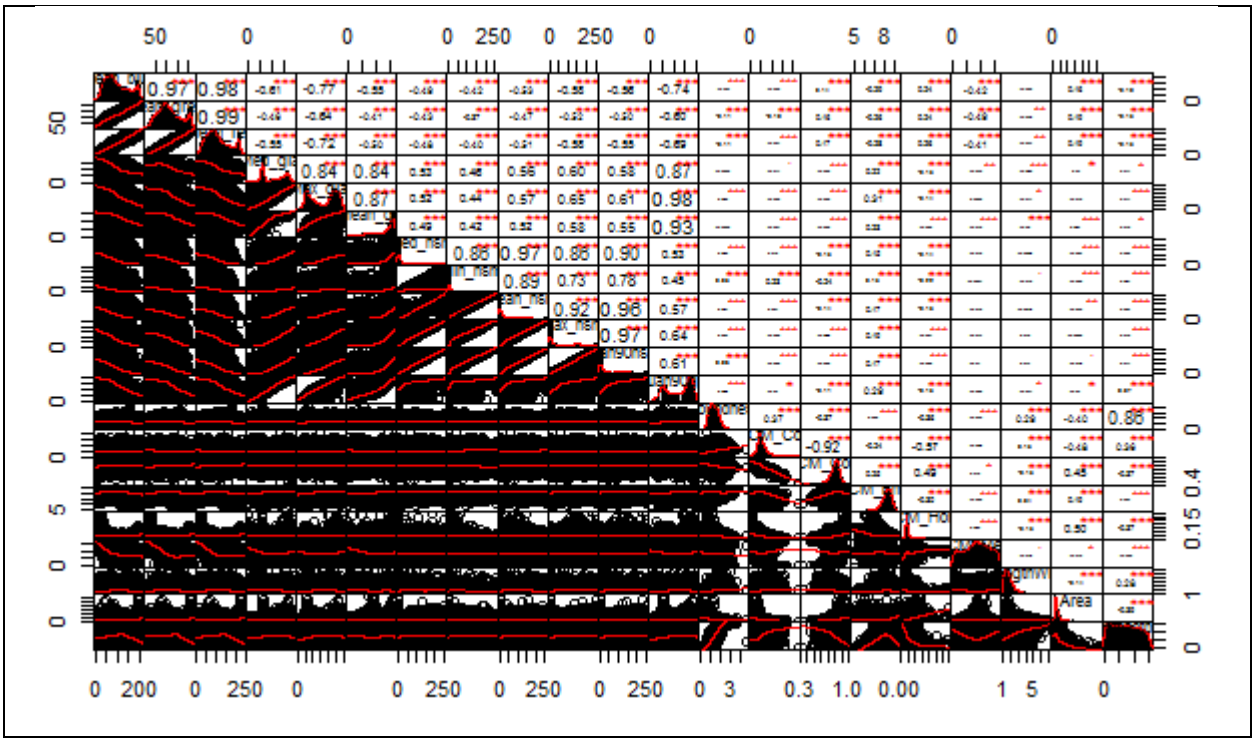

Figure 4.8. A correlation matrix for x01 with diagonal boxes showing each variable and their data distribution. Bottom left boxes = bi-variate scatterplots with mean line (red). Top right boxes = absolute correlation values. The diagonal variables in order from top left to bottom right are Mean_blue, Mean_green, Mean_red, Med_gla, Max_gla, Mean_gla, Med_nsm, Min_nsm, Mean_nsm, Max_nsm, quan90nsm, quan90gla, Roundness, GLCM_Contr, GLCM_Corre, GLCM_Entro, GLCM_Homog, GLCM_Mean_, LengthWidt, Area, and Compactnes.

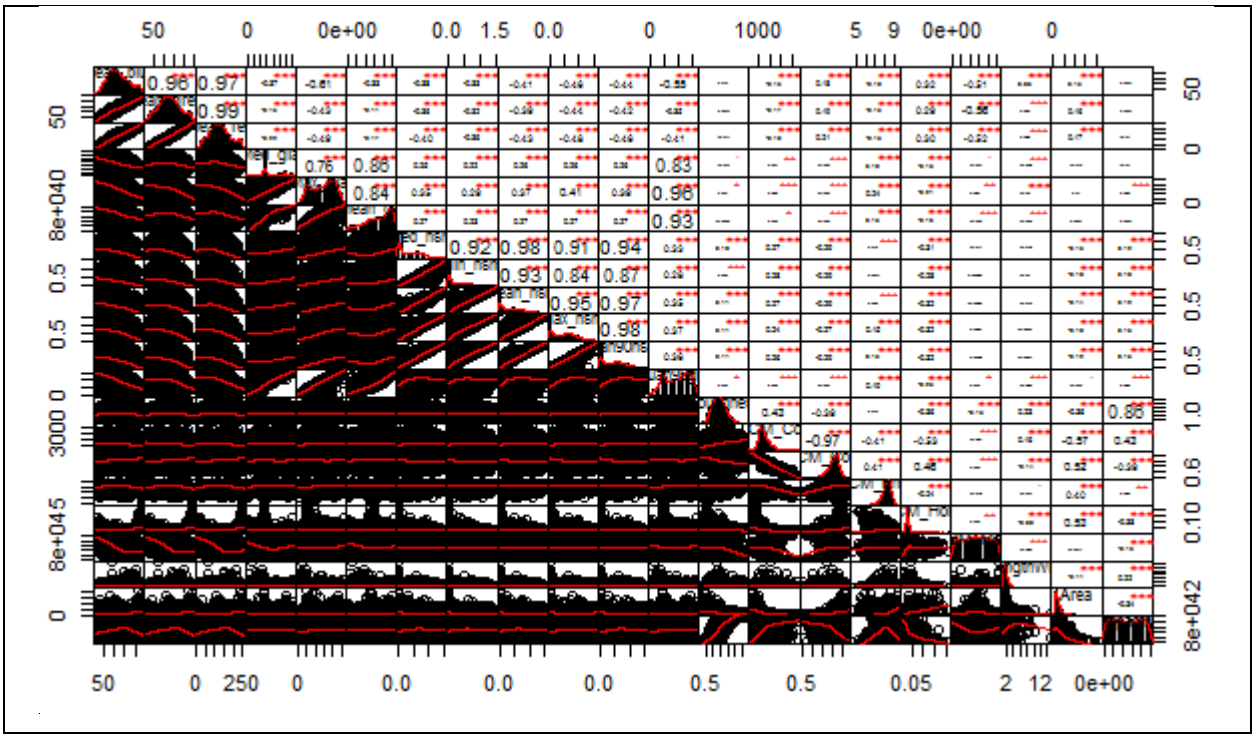

Figure 4.9. A correlation matrix for x06 with diagonal boxes showing each variable and their data distribution. Bottom left boxes = bi-variate scatterplots with mean line (red). Top right boxes = absolute correlation values. The diagonal variables in order from top left to bottom right are Mean_blue, Mean_green, Mean_red, Med_gla, Max_gla, Mean_gla, Med_nsm, Min_nsm, Mean_nsm, Max_nsm, quan90nsm, quan90gla, Roundness, GLCM_Contr, GLCM_Corre, GLCM_Entro, GLCM_Homog, GLCM_Mean_, LengthWidt, Area, and Compactnes.

Pearson correlation coefficient values between GLA variables returned high correlations for exclusion plot site 1. The Med_gla variable showed correlation values of 0.84, 0.84, and 0.87 with Max_gla, Mean_gla, and quan90gla respectively. Max_gla had a correlation of 0.87 with Mean_gla and 0.98 with quan90gla. Mean_gla showed a correlation value of 0.93 with quan90gla. The box plot assessment suggested Max_gla and Mean_gla to be the most dissimilar. However, these two variables returned an absolute correlation value which was greater than 0.6 and therefore did not satisfy assumption 1. Max_gla was deemed to have a slightly greater dissimilarity than Mean_gla. However, the Max_gla distribution appeared to be bimodal. This would not meet the assumption of normality required to meet assumption 2 of the MANOVA test. Both Mean_gla and Max_gla variables were retained for further x01 assumption testing.

Exclusion plot site 6 showed a similar trend in GLA correlations to exclusion plot site 1. Med_gla showed correlation values of 0.76, 0.88, and 0.83 with Max_gla, Mean_gla, and quan90gla respectively. Max_gla had a correlation of 0.84 with Mean_gla and 0.96 with quan90gla. Mean_gla showed a correlation value of 0.93 with quan90gla. Mean_gla and Max_gla were determined as the most dissimilar based on the box plot class distributions. Max_gla had the greatest dissimilarity when compared with Mean_gla. Max_gla was chosen for further assumption testing.

The height variables calculated were Med_nsm, Min_nsm, Mean_nsm, Max_nsm, and quan90nsm. These variables were strongly correlated to each other for both exclusion plot sites. The box plot analysis suggested that Med_nsm and Min_nsm for x01 and x06 may best differentiate between the classes on DHI. However, these variables returned a Pearson's correlation coefficient of 0.86 and 0.92 for x01 and x06 respectively. The data for Med_nsm and Min_nsm appeared to be skewed across both of the exclusion plot sites. To meet assumption 2 the data should be normally distributed. In this instance a transformation may need to be applied to mitigate the skew. However, transformations are not always successful. Med_nsm and Min_nsm for both exclusion plot sites were selected for further assumption testing. This will allow for the skew value to be quantified and to examine whether a transformation is applicable for the variables in order to satisfy the normality assumption of a MANOVA.

Texture was through the calculation of GLCM_Contr, GLCM_Corre, GLCM_Entro, GLCM_Homog, and GLCM_Mean_ variables. Only GLCM_Contr and GLCM_Corre showed a correlation value which did not satisfy assumption 1 for exclusion plot site 1. The box plot assessment showed that the variables GLCM_Entro and GLCM_Homog may provide the most suitable dissimilarity for classifying vegetation on exclusion plot site x01. Therefore, in the absence of an unsatisfactory correlation, GLCM_Entro and GLCM_Homog were selected for further assumption testing.

GLCM_Contr and GLCM_Corre also showed a high correlation of -0.97 for exclusion plot site x06. GLCM_Contr showed a slightly greater dissimilarity in the x06 box plot assessment and was recommended for further modelling. GLCM_Homog was also suggested as a probable variable to delineate x06 species and did not return any correlations. GLCM_Homog and GLCM_Contr may be the most suitable texture variables for classifying vegetation on exclusion plot site 6.

The final variable type was shape which included Roundness, LengthWidt, Area, and Compactnes. The correlation value of 0.86 was shown between Roundness and Compactnes for both exclusion plot sites. There were no other correlating shape variables. A value of 0.86 dissatisfies the assumption that correlations shouldn't be greater than 0.6 and therefore was not suitable for further MANOVA testing. The box plots for both x01 and x06 indicate that Compactnes and Area may differentiate between classes. Compactnes and Area for both exclusion plot sites were chosen for further modelling.

Correlations were predominately present between intra-variable types. However, there were a few inter-variable type correlations for exclusion plot site x01. Med_gla and Max_nsm had a Pearson correlation coefficient value of 0.6. To satisfy assumption 1 for a MANOVA the correlation value must be greater than 0.6. Therefore, Med_gla and Max_nsm would satisfy the assumption 1. Max_gla showed a correlation with Mean_green (-0.64), Max_nsm (0.65) and quan90nsm (0.61). Mean_blue also had high correlations with Med_gla and Max_gla displaying values of -0.61 and -0.77 respectively. Furthermore, quan90gla shared high correlation values with Mean_blue, Mean_green, Mean_red, Max_nsm, and quan90nsm. As Mean_blue, Mean_green, Med_gla, Max_nsm, and quan90gla were not selected for further modelling, these correlations did not impact the decision process.

Mean_red showed a high correlation with Max_gla (-0.72) for x01. From the GLA variables for x01 both Max_gla and Mean_gla were selected for further testing whilst Mean_red was the sole spectral reflectance variable selected for x01. Max_gla was removed from further assumption testing as GLA may still be modelled through the Mean_gla variable. The removal of Max_gla also mitigates potential issues arising from a bi-modal data distribution.

Exclusion plot site x06 showed less intra-variable correlations than x01. The correlation value for Mean_blue and Max_gla was -0.61. As Mean_blue was not selected for further modelling, this correlations did not impact the decision process.

4.5.2. Assumption 2 testing - normality

The histograms calculated in the correlation matrix provided information on whether the variables were normally distributed and thus satisfied assumption 2. However, further quantifiable analysis was required to eliminate human-induced bias. Skewness values were calculated for the refined

variable set (Table 4.2). Mean_gla and GLCM_Entro for x01 was reflected prior to transformation. A square root transformation was applied to Mean_red (x01), Mean_gla (x01), Med_nsm (x06), and Min_nsm (x06). The log of the data was derived for Med_nsm (x01), Min_nsm (x01), GLCM_Homog (x01 and x06), GLCM_Entro (x01), and GLCM_Contr (x06). Transformation outcomes may be found in Appendix 9.3.

Table 4.2. Skewness calculation per the variables selected for the MANOVA testing.

Variable	Skewness value a priori transformation		Skewness value posteriori transformation	
	x01	x06	x01	x06
Mean_red	0.52	0.25	0.05	-
Mean_gla	-0.86	-	0.28	-
Max_gla	-	-0.2	-	-
Med_nsm	1.13	0.61	0.22	-0.17
Min_nsm	1.64	0.98	0.6	0.19
GLCM_Homog	3.46	3.71	1.11	1.46
GLCM_Entro	-1.31	-	0.4	-
GLCM_Contr	-	1.23	-	0.03
Compactnes	0.1	0.02	-	-
Area	3.31	2.79	3.3	-0.12

After the transformations, a strong skew of greater than 1 was found in the variables GLCM_Homog for both exclusion plot sites and Area for x01. These variables were removed from subsequent modelling. Min_nsm for x01 showed a skewness value of 0.6 after the logarithmic transformations. As this value marginally surpasses the skewness threshold of 0.5 it was retained for further modelling.

There was an absence of outliers for both exclusion plot sites. Therefore, outliers would not affect the results of the study.

4.5.3. Updated correlation matrix and assumption 3 testing - linearity

The correlation matrix was re-run on the refined dataset and Med_nsm_TranF and Min_nsm_TranF for x01 returned a correlation of 0.88 (Figure 4.10). Furthermore, the bi-variate scatterplot mean line showed that the two variables are not linearly related. The transformed Min_nsm variable had shown that a numerical skewness was still present after the logarithmic transformation. From the two correlated transformed Normalised Surface Model (NSM) variables, Min_nsm_TranF was removed from further analysis due to the presence of a posteriori transformation skew.

Med_nsm_TranF and Min_nsm_TranF for x06 returned a correlation of 0.92 which is greater than the correlation assumption of less than 0.6 for assumption 2. Med_nsm_TranF was retained for further analysis due to the greater dissimilarity between classes shown in the box plot analysis. Min_nsm_TranF for x06 was excluded from the final MANOVA test.

Area_TranF and GLCM_Contr_TranF showed a strong correlation of -0.72 for x06. Due to difficulties determining box plot differentiation for Area_TranF, the transformed box-plots were re-analysed in Appendix 9.3. There was an increased dissimilarity found in GLCM_Contr_TranF when compared with Area_TranF. The latter variable was removed from the final x06 MANOVA test.

The x01 refined variable set appeared approximately linearly related once Min_nsm_TranF was removed, with the exception of GLCM_Entro_TranF and Compactness per the mean line of the bi-variate scatterplot. However, the non-linear relationship between GLCM_Entro_TranF and Compactness may still be suitable for a MANOVA as the strength of the collinearity between predictors is relatively low. Both the GLCM_Entro_TranF and Compactness for x01 were retained for the MANOVA test.

The refined variable set for x06 appeared approximately linear across all relationships. No further variable exclusion was required.

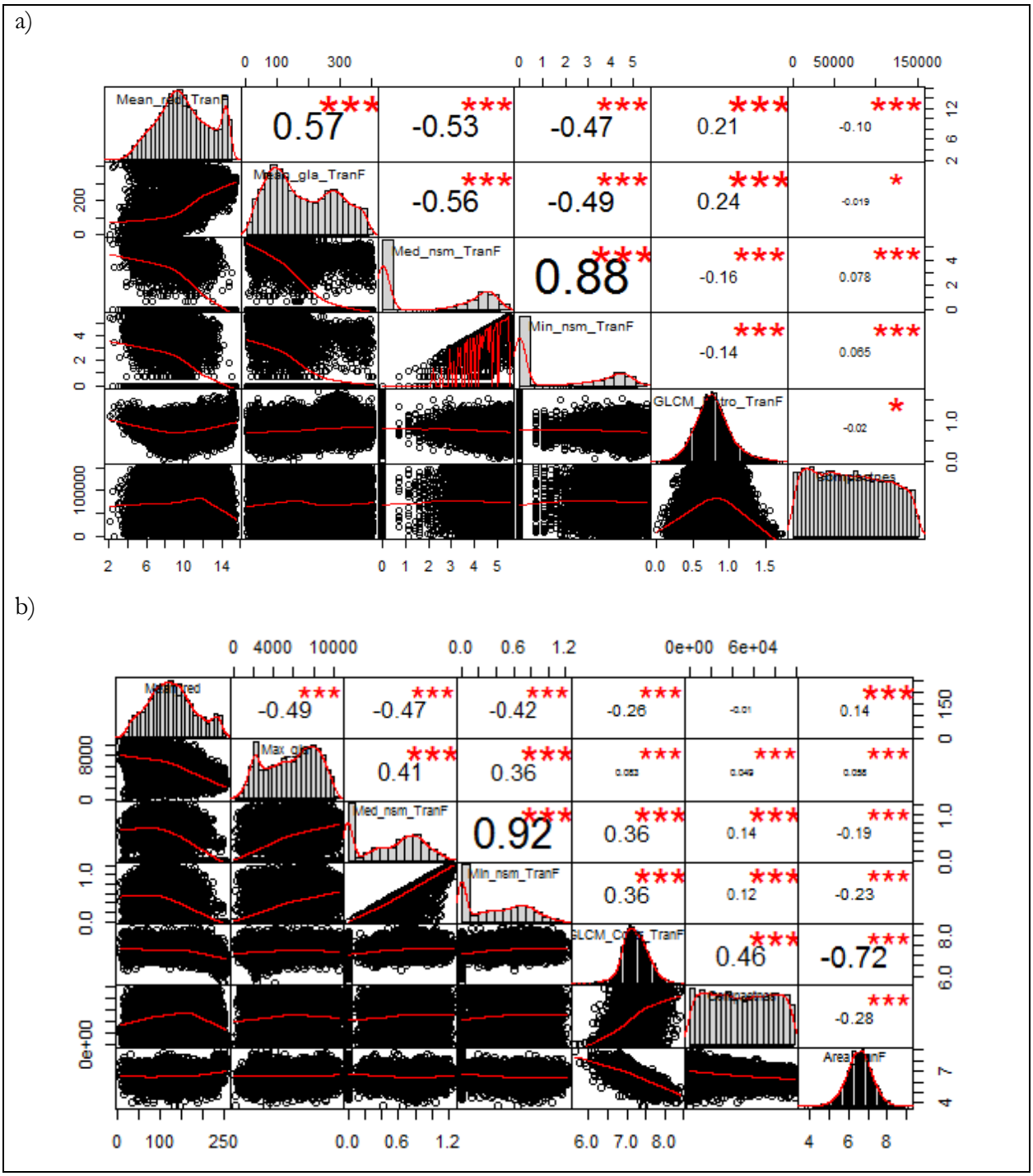

Figure 4.10. Final correlation matrices with diagonal boxes showing each variable and their data distribution. a) x01 and b) x06. Bottom left boxes = bi-variate scatterplots with mean line (red). Top right boxes = absolute correlation values.

4.5.4. Multi-variate Analysis of Variance and post-hoc univariate test

The final dataset derived for x01 contained the dependent (Species) and refined independent variable set after assumption testing. Predictor variables for testing significance were Mean_red_TranF, Mean_gla_TranF, Med_nsm_TranF, GLCM_Entro_TranF, and Compactnes. The MANOVA test showed that the variable set selected had a significant effect on Species (Table 4.3). A pairwise comparison suggests that all predictor variables are statistically significant for delineating species on DHI (Table 4.4).

Table 4.3. Results of the MANOVA applied to the effect of the segment variables on Species for x01. Significance was measured at a $P < 0.05$ level and indicated by *.

	df	Pillai test	P	P < 0.05
Species	9	1.12	<0.001	*
Residuals	13,648			

Table 4.4. Result of the Summarise Analysis of Variance test for the effect of x01 segment variables on delineating species. Sum Sq = Sum of Squares. Mean Sq = Mean sum of Squares. Significance was measured at a $P < 0.05$ level and indicated by *.

	Sum Sq	Mean Sq	P	P < 0.05
Mean_red_TranF	62,240	6,915.6	<0.001	*
Mean_gla_TranF	95,692,547	10,632,505	<0.001	*
Med_nsm_TranF	33,736	3,748.4	<0.001	*
GLCM_Entro_TranF	104.68	11.63	<0.001	*
Compactnes	1.53e+12	1.69e+11	<0.001	*

Significance testing for x06 was conducted on Mean_red, Max_gla, Med_nsm_TranF, GLCM_Contr_TranF, and Compactnes. The MANOVA test showed that the variable set selected had a significant effect on Species (Table 4.5). A pairwise comparison suggests that all predictor variables are statistically significant for delineating species on DHI (Table 4.6).

Table 4.5. Results of the MANOVA applied to the effect of the segment variables on Species for x06. Significance was measured at a P < 0.05 level and indicated by *.

	df	Pillai test	P	P < 0.05
Species	4	0.82	<0.001	*
Residuals	5,952			

Table 4.6. Result of the Summarise Analysis of Variance test for the effect of x06 segment variables on delineating species. Sum Sq = Sum of Squares. Mean Sq = Mean sum of Squares. Significance was measured at a P < 0.05 level and indicated by *.

	Sum Sq	Mean Sq	P	P < 0.05
Mean_red	8,232,932	2,058,233	<0.001	*
Max_gla	1.39e+10	3.47e+9	<0.001	*
Med_nsm_TranF	424.97	106.24	<0.001	*
GLCM_Contr_TranF	118.01	29.5	<0.001	*
Compactnes	1.91e+11	4.76e+10	<0.001	*

4.6. Random forest classification accuracy

Accuracy outcomes measured showed that a scale parameter set at 20 provided the greatest classification accuracy (Kappa = 0.7, accuracy = 0.75) for exclusion plot site x01 followed closely by a scale parameter of 50 tested with all variables (Kappa = 0.69, accuracy = 0.74) (Table 4.7). All variables in this instance denotes all originally captured variables and not the transformed variables. A scale value of 20 for x01 had a higher computation time (>1hr) compared to a scale value of 50 tested with all variables (5.5mins). The difference between accuracy outcomes is marginal for the scale of 20 compared to 50 with all variables and the computation time is markedly higher.

Both exclusion plot sites showed the segmentation scale of 100 as having a lower Kappa value and accuracy estimate than the scenarios where the scale was calculated at 20 and 50 with all variables. This suggests there may be a decrease in classification accuracy when the segments are more generalised. Therefore, a scale of 100 for both exclusion plot sites may not be suitable for quantifying species on DHI.

Table 4.7. Random Forest out of bag Kappa and accuracy quantified by mean value over 10 iterations per scenario. Approximate computation time captured from a random iteration sample per scenario.

Scenario	Kappa	Accuracy	Computation time
Exclusion plot site x01			
Scale 20 – all variables	0.7	0.75	>1 hour
Scale 50 – untransformed selected variables	0.5	0.59	2 minutes
Scale 50 – transformed selected variables	0.51	0.59	2 minutes
Scale 50 – all variables	0.69	0.74	5.5 minutes
Scale 100 – all variables	0.63	0.69	4 minutes
Exclusion plot site x06			
Scale 20 – all variables	0.7	0.78	45 minutes
Scale 50 – untransformed selected variables	0.56	0.63	1.75 minutes
Scale 50 – transformed selected variables	0.51	0.63	1.75 minutes
Scale 50 – all variables	0.67	0.74	5.5 minutes
Scale 100 – all variables	0.57	0.66	3.5 minutes

The MANOVA dimension reduction of variables for a scale of 50 resulted in a classification accuracy decrease for exclusion plot site 1 from 74% to 59% for the all variables and untransformed selected variable scenarios respectively. However, computational performance was improved by 63.64%.

A scale of 50 with all variables may provide the most efficient metric for classifying exclusion plot site x01 segmentation outputs. Values for all scenario iterations across both exclusion plot sites may be found in Appendix 9.4.

The highest accuracy outcome for exclusion plot site 6 was the scale of 20 followed by 50 tested with all variables. There was a 4% classification accuracy difference between these scenarios. The scenarios returned similar Kappa values of 0.7 and 0.67 for a scale of 20 and 50 tested with all

variables respectively. The Kappa value is an equation which tests how far the data classification departs from expected outcomes. A higher Kappa value suggests the classification is less likely to be calculated by chance. The Kappa value comparison between x06 scale parameterisation of 20 and 50 with all variables suggests both scenarios may provide a suitable classification. The scale of 50 with all variables had a lower computation time (5.5minutes) when compared with a scale of 20 (45 minutes) for x06.

There was no significant difference between the transformed and untransformed refined variable set per a scale of 50 for x06. The x06 segment scale of 50 untransformed variable subset reduced the accuracy and Kappa values when compared with a scale of 50 tested using all variables.

The MANOVA dimension reduction of variables for a scale of 50 resulted in a classification accuracy decrease for exclusion plot site 6 with 74% to 63% for the all variables and untransformed selected variable scenarios respectively. However, computational performance was improved by 68.18%.

The segmentation scale parameter of 50 calculated with all variables for exclusion plot site 6 was selected for the final classification output and class comparisons due to computational performance outcompeting the similarly accurate segment scale of 20 scenario.

Exclusion plot site 1 returned a slightly lower classification accuracy for all scenarios when compared with exclusion plot site 6. This may be related to the two identified species found across x01 which were removed due to having less than 10 samples each. Classification accuracy may be improved if an exhaustive survey effort is applied. Achieving a minimum of 10 samples for all classes when in the field may improve accuracy outcome in subsequent modelling stages.

4.7. Area comparisons

Class comparisons were made between 15x15m quadrats, one within the exclusion plot and another mirrored external to the aforementioned quadrat for x01 and x06 (Figure 4.11-12, Table 4.8-9).

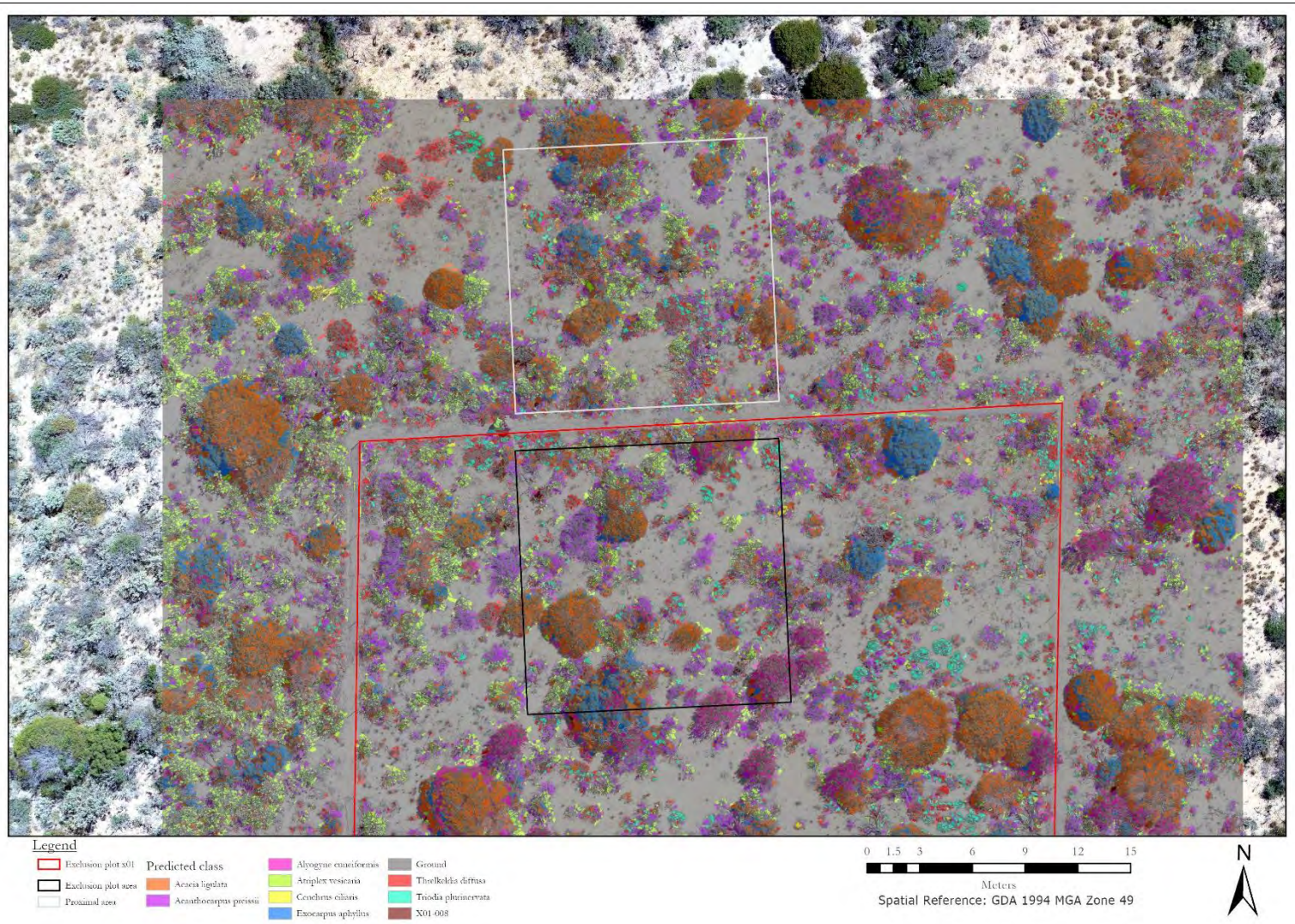

Figure 4.11. Northernmost exclusion plot site (x01) map with predicted segments and remotely piloted aircraft capture. Red polygon = exclusion plot. Black polygon = 15x15m quadrat within the exclusion plot. Grey polygon = 15x15m quadrat located in mirrored proximal area to the quadrat measured within the exclusion plot. Final iteration to calculate this prediction quantified Kappa value and accuracy as 0.69 and 74% respectively.

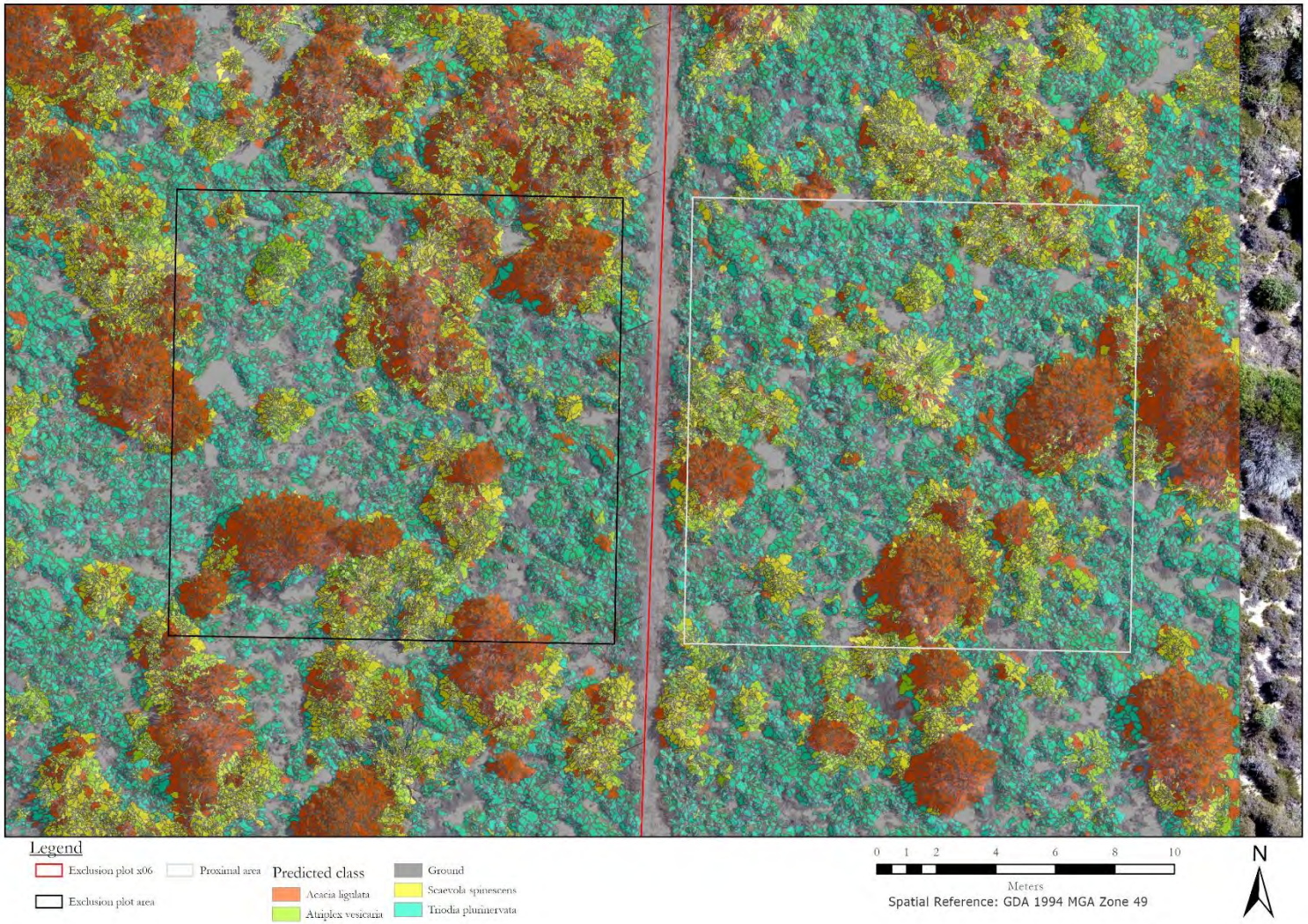

Figure 4.12. Southernmost exclusion plot site (x06) map with predicted segments and remotely piloted aircraft capture. Red polygon = exclusion plot. Black polygon = 15x15m quadrat within the exclusion plot. Grey polygon = 15x15m quadrat located in mirrored proximal area to the quadrat measured within the exclusion plot. Final iteration to calculate this prediction quantified Kappa value and accuracy as 0.67 and 74% respectively.

Table 4.8. Area comparisons quantified per pixel area and % for exclusion plot x01 testing quadrats.

Class	Area within exclusion plot (area per pixel)	Proximal area (area per pixel)	Area within exclusion plot (%)	Proximal area (%)	Change (%)
Acacia ligulata	1,207,440	1,291,156	15.69	16.76	+1.07
Acanthocarpus preissii	1,178,146	1,132,710	15.31	14.7	-0.61
Alyogyne cuneiformis	493,440	235073	6.41	3.05	-3.36
Atriplex vesicaria	903,687	1,043,812	11.74	13.55	+1.81
Cenchrus ciliaris	28,041	65,017	0.36	0.84	+0.48
Exocarpus aphyllus	246,243	308,211	3.2	4	+0.8
Ground	2,818,084	2,749,670	36.62	35.69	-0.93
Threlkeldia diffusa	522,952	541,302	6.8	7.03	+0.23
Triodia plurinervata	244,852	250,426	3.18	3.25	+0.07
X01-008	52,365	87,196	0.68	1.31	+0.63

Table 4.9. Area comparisons quantified per pixel area and % for exclusion plot x06 testing quadrats.

Class	Area within exclusion plot (area per pixel)	Proximal area (area per pixel)	Area within exclusion plot (%)	Proximal area (%)	Change (%)
Acacia ligulata	1,308,736	985,756	20.41	15.36	-5.05
Atriplex vesicaria	456,052	715,825	7.11	11.16	+4.05
Ground	968,523	705,660	15.11	11	-4.11
Scaevola spinescens	1,043,865	892,798	16.28	13.91	-2.37
Triodia plurinervata	2,633,579	3,116,428	41.08	48.57	+7.49

There was a minimal difference between class coverage for x01. The highest differential was shown in the *A. cuneiformis* species with a -3.36% change. *A. vesicaria* was the second highest outcome with an increase of 1.81% cover between the quadrat within the exclusion plot and the proximal area. *T. plurinervata* showed the least change (+0.07%) between x01 quadrats.

A. ligulata showed to have the greatest coverage of the plant species both within the exclusion plot and across the external quadrat for x01. A visual assessment showed that *A. ligulata* could be found sharing habitat with similarly sized shrubs *E. aphyllus* and *A. preissii*.

Comparisons between 15x15m quadrats for x06 showed an increase in *T. plurinervata* (7.49%) and *A. vesicaria* (4.05%) cover (Figure 4.12, Table 4.9). There was a subsequent decrease in *A. ligulata*, ground, and *S, spinescens* classes with values of 5.05%, 4.11%, and 2.37% respectively for x06.

T. plurinervata followed by *A. ligulata* showed to have the greatest coverage of the plant species both within the exclusion plot and across the external quadrat for x06. A visual assessment showed that the lower-lying forb *S. spinescens* and grass *T. plurinervata* could be found growing peripherally to *A. ligulata*.

Ground coverage found at x01 was greater than the area of ground found for x06. The increase in diversity and ground may relate to a decrease in *A. ligulata* cover for x01 when compared to x06.

Using a segmentation method which divides objects into multiple segments has shown to delineate the inherent complexities of wild plant growth for both exclusion plot sites.

5. Discussion

Object-based classification techniques on very high-resolution aerial imagery has been shown to be a viable method of delineating vegetation to a species level. Comparisons between the experimental control quadrats and proximal quadrats tested the impacts of translocated fauna posteriori ungulate removal. This section will provide an in-depth discussion of the five objectives for this report: a) construct and test a framework of sampling vegetation for the purpose of classifying high-resolution aerial imagery, b) segmentation of aerial imagery and implementation of the Random Forest (RF) algorithm to classify segmented data, c) dimension reduction to obtain variables offering the greatest degree of species level separation, d) accuracy assessment of the RF classifier, and e) determining if translocated fauna are having an ecosystem engineered impact on Dirk Hartog Island (DHI).

5.1. Data capture and processing

Incidental surveying of vegetation is where the sample is found by chance and the represented target is not based on a pre-defined taxa (Yong et al., 2018). Sampling of the incidental specimen may then be conducted based on a defined numerical threshold i.e. the aim of capturing 10 samples per species for this report. The application of this observation technique on DHI was found to be a proficient sampling method. Sampling was conducted in an area relative to the exclusion plot site. Segment classification may have been interpolated for the entire exclusion plot site Remotely Piloted Aircraft (RPA) capture. However, there may be a species bias if sampling coverage is restricted to one section of the capture. Where the species richness was greater (x01), the species sampling effort took a longer period of time to cover the same area than the exclusion plot site where the species richness was less (x06). Two recommended alternatives may improve the field survey sample distributions. Firstly, to sample every second or third incidental specimen. This would allow for a greater area to be sampled whilst maintaining approximately the same level of survey effort.

An alternative is to divide the current segmentation boundary areas (x01 = 4,215.1m^2 and x06 = 3,000.66m^2) into four plots. These boundaries determine the minimum area required to capture the species diversity. Tobler's first law of geography states that phenomena nearby to each other are more closely related than phenomena further away (Tobler, 1970). Therefore, distributing observed classes within the study design would allow for segments located further from the exclusion plot boundary

to also be accurately classified. Dispersing these plots across the RPA imagery would allow for a more distributed sampling effort and retain the same level of effort.

Segmentation totals for *Acacia ligulata* and *Alyogyne cuneiformis* presented the highest count for x01. However, *A. ligulata* and *A. cuneiformis* showed the second most number of field samples behind *Atriplex vesicaria*. *A. ligulata* segments for x06 also presented the highest number of total segments with *A. vesicaria* returning under half the segments calculated than that of the *A. ligulata*. *A. ligulata* was sampled an approximate equivalent number of times when compared to *A. vesicaria* across x06. *A. cuneiformis* was not found at x06. *A. vesicaria* is a medium sized bush which may grow between 0.3-1m tall (Spooner, 1999b). *A. ligulata* and *A. cuneiformis* are both larger structures both shown to have a maximal growth of 4(-6 in some environments) and 3.5m respectively (Paczkowska, 1996; Spooner, 1999a). Vegetation classes which experience larger growth patterns may have a higher segmentation count due to an increase in canopy. Therefore, the discrepancy seen in field capture sample size and total number of segments may be explained by differences in growth behaviours.

Cenchrus ciliaris and *Triodia plurinervata* are both tussock-forming graminoids which are relatively low lying with a maximum growth of up to 1.5m, allowing for successful seed wind dispersal (Coleman, 1998; Paczkowska, 1993). This study found that *C. ciliaris* showed the lowest segment total for x01 whilst *T. plurinervata* returned the lowest segmentation count at x06. However, *C. ciliaris* placed within the upper half of overall field samples based on the plant species and ground class included in the modelling process for x01. Furthermore, the number of field samples for *Cenchrus ciliaris* was relative to the other classes for x06. The reduction in the segment count compared with other larger growth forms may suggest a correlation between plant morphology and segmentation total.

Moderately sized plant species for x01 include *Atriplex vesicaria* and *Exocarpus aphyllus* shrubs and the forb *Acanthocarpus preissii* (Paczkowska, 1994, 1995a; Spooner, 1999b). The medium plants found at x01 were centrally ranged with segmentation totaling 903-1,637. *Atriplex vesicaria* and *Scaevola spinescens* represent moderately sized sample for x06 with a total of 790 and 1,467 segments respectively (Spooner, 1999b, 2000). There is a difficulty in placing the moderately size plant species within a range for x06 due to a decrease in species richness when compared to x01. Overall, the differences in segmentation size against plant species morphological characteristics are more apparent in x01. The increase in species richness may provide a greater understanding of the plant type to segmentation

total relationship. Further studies and required to understand the relationship between sample size, segmentation count and plant growth behaviours. This will allow environmental managers to determine the most efficient sampling method.

5.2. Segment parameterisation

The unit of interest in this study were plant species. Each individual was divided into multiple segments to increase the sample size and therefore the strength of classification (Millard & Richardson, 2015). The unique characteristics of the individual plant segment mosaics allowed for their varying exhibited qualities to be captured. For example, individuals which have had the leaf density reduced on one side due to wind erosion or where two plants are sharing the same ground area for growth.

Yang et al. (2019) suggest that dividing a feature into multiple segments is the result of over-segmentation and is inadvisable. However, their study measured urban areas and crop types from Sentinal-2A and Landsat 8 satellites at a regional scale (Yang, Mansaray, Huang, & Wang, 2019). This would not be applicable to a species-level classification of very high-resolution RPA data. Each measured surface contains unique qualities which may result in different parameterisation requirements (Singh et al., 2005). The outputs for this study highlighted peripheral grasses and small forbs which were growing underneath larger shrubs. Another example shown in the classified outputs were instances where two large shrubs growing in the same area could be discriminated. Leveraging the power of segment clusters has shown to capture the complexity of vegetation on DHI which may not be possible if plants were modelled as a whole.

A visual assessment of the scale value to determine if segmentation represented the underlying aerial imagery was shown to reliably validate parameterisation. This method is in line with the broader scientific literature (Adam et al., 2016; Juniati & Arrofiqoh, 2017; Mathieu et al., 2007; Myint et al., 2011; Sertel & Alganci, 2015; Sibaruddin et al., 2018; Yan et al., 2007; Yu et al., 2006).

This research identified a probable linear relationship between segmentation scale and the accuracy of the classification. As the objects of interest became more segmented, the out-of-bag estimates improved. This may be related to how well the segments fit the aerial image. Segmentation quality is correlated with the image object geometry and has been shown to influence the classification accuracy (Benz et al., 2004; Niphadkar et al., 2017; Zhang et al., 2012). However, reliance of accuracy

assessment calculations may result in over- or under-segmentation (Jing et al., 2012; Kawakubo et al., 2013). The visual assessment of the segmentation output shown in this research paper may eliminate difficulties in determining whether accuracy outcomes relating to parameterisation are suitable.

This study showed that scales of 20, 50, and 100 all may potentially represent individual plant species and the surrounding surface area. Identifying whether a linear relationship exists between accuracy and scale values for vegetation on DHI may improve classification performance for future assessments. This may provide resource managers a more precise delineation of flora for long-term monitoring programs.

5.3. Variable selection

Application of all multi-spectral and ancillary variables was shown to outcompete the refined dataset in terms of classification accuracy (15% and 11% reduction for x01 and x06 respectively). Studies have shown that the Green Leaf Algorithm (Booth et al., 2005; Chianucci et al., 2016; Louhaichi et al., 2001; Macfarlane & Ogden, 2012), Normalised Surface Model (Booth et al., 2005; Kluckner et al., 2009; Louhaichi et al., 2001; Räsänen et al., 2014; Salehi et al., 2012), Grey-Level Co-occurrence Matrix (Champion et al., 2013; Feng et al., 2015; Kim et al., 2011; Liu et al., 2013; Wang et al., 2004), and segment geometry (Jiao & Liu, 2012; Mathieu et al., 2007; Memarian et al., 2013) may compliment red, green, and blue bands in an object-based classification of remotely sensed data. The breadth of variables used in this report may functionally exaggerate the spectral characteristics of certain plant species or complement each other to effectively discriminate vegetation types and the ground class. However, the use of box-plot comparisons and assumption testing methods achieved a strict dimensionality reduction of 76% for both exclusion plot sites. Confining the variable selection to a box-plot analysis and subsequent correlation matrix may provide an equivalent accuracy output to the all variables scenario whilst mitigating the 'curse of dimensionality'.

The 'curse of dimensionality' phenomena is where there is a relationship between the number of variables input into the algorithm and the temporal performance of the algorithm (Jin & Liu, 2010). Therefore, 'curse of dimensionality' mitigation methods have been proposed to reduce the number of features input by eliminating redundant variables to improve computational performance (Camps-Valls, Tuia, Gómez-Chova, Jiménez, & Malo, 2011; Witten, Frank, & Hall, 2011). The three methods

for dimensionality reduction are filter, wrapper, and embedded (Camps-Valls et al., 2011; Kavzoglu & Mather, 2002).

Wrapper dimensionality reduction techniques incorporate classification algorithms to evaluate the strength of a subset's usefulness in relation to a predictor (Talavera, 2005). These include Support Vector Machines and Greedy Search Algorithms (Guyon & Elisseeff, 2003). The benefit of these method is that relationships between classes are validated (Camps-Valls et al., 2011). However, cross-validation of variables and classes is timely and may exceed the computational performance of the filter method or the scenario of processing all variables (Colkesen & Kavzoglu, 2018). This renders the purpose of utilising a wrapper variable reduction method void. Furthermore, the wrapper methods may lend themselves to overfitting, especially where there is a low sample size (Colkesen & Kavzoglu, 2018).

Embedded methods select the variable subset during the machine learning algorithm training process (Camps-Valls et al., 2011). For example, Puissant, Rougier, and Stumpf (2014) used repeated Random Forest variable importance outcomes to determine variables that may be removed from further classification. For each iteration the least significant features (lower 20%) were removed. Their study reduced the initial dataset from 80 to 33 variables (Puissant, Rougier, & Stumpf, 2014). However, this process is counter-intuitive if the rationale of dimension reduction is to minimise the number of inputs prior to machine learning computation. Furthermore, Yang et al. (2019) suggest that variable importance methods may not interpret the interplay between variable types. A lowly ranked feature may extricate additional dissimilarities when complemented with other variables (Yang et al., 2019). Ignoring the relationships between variables may result in a loss of differentiation capability.

Lastly, the filter method is used to subset the data based on indirect measures prior to classification e.g. Principal Component Analysis (PCA), Nonnegative Matrix Factorization (NMF), Chi-squared testing, and the process delineated for this research paper (Colkesen & Kavzoglu, 2018; Huang & Zhang, 2008). Studies may also fuse filter methods to obtain a reduced dimensionality and/or apply the technique to a sub-group rather than the entire dataset (Abasi, Arefi, & Bigdeli, 2015; Chen et al., 2018).

Chen et al. (2018) used a PCA to reduce the original 32 texture variables to 12. The texture variable subset were amalgamated with spectral reflectance, terrain, and spatial relation features. Their object-based classification of GF-1 satellite imagery achieved a maximum classification accuracy of 88.16% using Random Forest. The application of PCA to a sub-group was required because PCA is a holistic algorithm (Huang & Zhang, 2008). PCA, Chi-squared, and NMF filter methods may not have an intuitive interpretation of the relationship between classes. A box-plot analysis allows for the recognition of intra-class relations. The proposed flow-on correlation matrix may test sub-grouping dissimilarities and holistic relationships.

Limiting the variable refinement to the initial box plot assessment or correlation assumption test, as done in this study, may provide an effective method for reducing the variable set for computational efficiency whilst retaining a satisfactory accuracy. A trade-off in this instance may be made where a slight reduction in accuracy is accepted in return for expedient data processing. This may allow for discriminatory pre-selection of variables in future research projects. Further studies to test the application of solely the box-plot and correlation matrix dimensionality reduction are required. This would determine if refinement for computational performance improvement may be achieved using this method with an equivalent accuracy to the application of all variables.

5.4. Accuracy assessment

The out-of bag estimates produced by RF determines the accuracy of the classification through the use of bootstrapped samples cross-validating outcomes (Breiman, 2001). However, the random forest accuracy assessment was developed for testing two-dimensional datasets. Spatially dependent data may result in RF accuracy estimates which are auto-correlated (Haining, 2015). With multiple segments per plant, there is a chance that the accuracy calculation may be influenced by randomly sampling a cluster. RF bootstrapped samples do not factor in the spatial relationships of the data being inputted (Breiman, 2001).

Independent assessments of error may offer a more reliable evaluation method for geospatial datasets (Millard & Richardson, 2015). A confusion or error matrix visually determines the performance of a classifier through the calculation of the overall accuracy, users accuracy (false positive: type 1 error), and producers accuracy (false negative: type 2 error) (Congalton & Green, 2009; Lillesand et al., 2015). The user may sample one or more observed segment per sample to ensure the

data is dispersed. Cross-tabulation of classes has been shown to reliably assess the accuracy of object-based classification (Chapman et al., 2010; Duro et al., 2012; Jhonnerie et al., 2015; Ye et al., 2018).

The difficulty with an independent error assessment is that take additional time to program and analyse in comparison with accuracy assessments built into the model. Due to the volume of segments (165,935 for x01 and 104,800 for x06) the effects of spatial auto-correlation on accuracy outcomes are likely to be minimal. Therefore, it is discretionary whether an independent error matrix is applied for future object-based modelling of vegetation on DHI.

5.5. Ecosystem engineering

Species class cover between 15 x 15 m quadrats within the exclusion plot and external to the exclusion plot for x01 showed a negligible difference. However, there was a small scale change in species level canopy coverage across x06 for *T. plurinervata* (+7.49%), *A. vesicaria* (+4.05%) cover, *A. ligulata* (-5.05%), ground (-4.11%), and *S. spinescens* (-2.37%) classes. Further temporal studies may show if these differences relate to the landscape providing more favourable habitat for the translocated endangered species to DHI.

Rufous and banded hare-wallabies are known to create nests and dig small burrows within *T. plurinervata* and *A. ligulata* respectively (Burbidge & Woinarski, 2016a, 2016b; Cowen et al., 2018). The rufous hare wallaby may also utilise *A. ligulata* for habitat (Burbidge & Woinarski, 2016a). A site which is abundant in *T. plurinervata* and *A. ligulata* could provide ideal habitat for either species. Exclusion plot site 6 showed a greater composition area of *A. ligulata* and *T. plurinervata* than exclusion plot site 1.

Bare ground, seedlings, and juveniles coverage found at x01 was greater than the area of the 'ground' class for x06. The reduction of open space and subsequent increase in plant density may provide an ideal refugium for the risk averse hare-wallabies (Hardman & Moro, 2006; Richards, 2012). Hardman and Moro (2006) studied mala (*Lagostrophus fasciatus* and *Lagorchestes hirsutus*) habitat preferences using animals fitted with radio-transmitters for diurnal monitoring on a protected area of the Peron Peninsula. They found that both species sought out a floristically and structurally variable vegetation outside of their known preferences due to the former providing a greater habitat density and thus reliable refugia. Exclusion plot sites that combine favourable flora species with vegetation

density may be more preferable for habitation leading to an increase in changes to the species composition.

Interestingly, *T. plurinervata* increased in cover whilst *A. ligulata* decreased. Ecosystem engineering through digging behaviours can promote plant growth in adjacent soil heaps and lower conductivity, phosphorus, potassium, sulphur, and microbial activity of direct foraging pits (Valentine et al., 2018). Testing direct hare-wallaby nesting within *A. ligulata*, in conjunction with remotely sensed data may show that biopedturbation is promoting plant growth of peripheral grasses such as the *T. plurinervata*.

The absence of changes for x01 may be the result of ecosystem engineering inertia. Studies have shown a temporal lag between organic litter build up and altered abiotic conditions for microhabitats and the creation of beneficial growing conditions (Eldridge & James, 2009; James et al., 2009; Mallen-Cooper et al., 2019; Sandom et al., 2013; Travers et al., 2012; Valentine et al., 2016). Ecosystem engineering inertia to providing beneficial growing patterns may also explain the minimal differences found for x06. The introduced hare-wallaby species have been present on DHI for less than 2 years. Long-term monitoring of exclusion plot sites is required to determine if translocated hare-wallabies are having an ecosystem engineered impact.

Finally, the difference shown in exclusion plot site 6 may relate to natural variation across the landscape and not faunal biopedturbation. Environmental stochasticity is known pillar of natural systems (Fujiwara & Takada, 2017). Further studies utilising the entirety of the 40 x 40 m exclusion plot would reduce uncertainties associated with whether the vegetation mosaic difference has been calculated by chance or significantly impacted by translocated hare-wallabies on DHI.

6. Conclusions

Object-based classification of very high-resolution Remotely Piloted Aircraft (RPA) imagery has shown to be an effective method for delineating vegetation on Dirk Hartog Island (DHI) to a species level. This method may be viable for replication when assessing temporal changes in vegetation on DHI. It is recommended that long-term field monitoring programs implement a distributed stratified, systematic random sampling approach in the form of sampling every second/third incidental plant or dividing the survey effort into four dispersed quadrats. This may provide an equally accurate classification of the broader RPA capture area. Further studies measuring the relationship between plant growth behaviours, sample size per species, and segmentation total may also provide increased survey effort efficacy.

Segment parameterisation was found to have a potential association with the calculated out-of-bag accuracy and Kappa values. Research testing this relationship using high-resolution aerial imagery captured on DHI may improve segment parameterisation and classification accuracy for future monitoring of plant species and faunal impacts.

Implementation of Multi-variate Analysis of Variance assumption and post-hoc statistical testing showed to be a strict dimensionality reduction method. A trade-off was made between improving computational performance through decreasing variables and the classification accuracy. Variable refinement may be restricted to solely the box plot analysis and correlation matrix to preserve slightly more information. The outcome of a less strict reduction may extract variables which show dissimilarities between classes whilst retaining an acceptable accuracy outcome.

There was minimal differences found between both the exclusion plot sites and their associated quadrat comparatives. Long-term vegetation monitoring is recommended to adequately determine the impacts of translocated fauna. Furthermore, utilising the entirety of RPA capture may reduce uncertainties to whether differences are related to environmental stochasticity or faunal impacts. Data captured and analysed within the research paper may provide base-line composition statistics for future object-based modelling of DHI.

7. Bibliography

Abasi, B., Arefi, H., & Bigdeli, B. (2015). *A dimension reduction-based method for classification of hyperspectral and LiDAR data*. Paper presented at the International Conference on Sensors & Models in Remote Sensing and Photogrammetry, Kish Island, Iran.

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*(10), 3440-3458. doi:https://doi.org/10.1080/01431161.2014.903435

Adam, H. E., Csaplovics, E., & Elhaja, M. E. (2016). *A comparison of pixel-based and object-based approaches for land use land cover classification in semi-arid areas, Sudan* Paper presented at the 8th IGRSM International Conference and Exhibition on Remote Sensing & GIS, Kuala Lumpur, Malaysia.

Addink, E. A., Jong, S. M. D., & Pebesma, E. (2007). The importance of scale in object-based mapping of vegetation parameters with hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing, 73*(8), 905-912. doi:https://doi.org/10.14358/PERS.73.8.905

Agisoft. (2018). Photoscan-pro (Version 1.4.2). St Petersburg, Russia: Agisoft LLC. Retrieved from https://www.agisoft.com

Akar, Ö., & Güngör, O. (2015). Integrating multiple texture methods and NDVI to the Random Forest classification algorithm to detect tea and hazelnut plantation areas in northeast Turkey. *International Journal of Remote Sensing, 36*(2), 442-464. doi:https://doi.org/10.1080/01431161.2014.995276

Alberto, R. T., Serrano, S. C., Damian, G. B., Camaso, E. E., Celestino, A. B., Hernando, P. J. C., . . . Tagaca, R. C. (2016). *Object based agricultural land cover classification map of shadowed areas from aerial image and LiDAR data using Support Vector Machine*. Paper presented at the XXIII ISPRS Congress, Prague, Czech Republic.

Ali, M. Z., Qazi, W., & Aslam, N. (2018). A comparative study of ALOS-2 PALSAR and landsat-8 imagery for landcover classification using maximum likelihood classifier. *The Egyptian Journal of Remote Sensing and Space Sciences, 21*, s29-s35. doi:https://doi.org/10.1016/j.ejrs.2018.03.003

Arbiol, R., Zhang, Y., & Palà, V. (2006). *Advanced Classification Techniques: A Review*. Paper presented at the ISPRS Commission VII Mid-term Symposium "Remote Sensing: From Pixels to Processe, Enschede, Netherlands.

Asher, J., & Morris, K. (2015). *Dirk Hartog Island biosecurity implementation plan - a shared responsibility*. Perth, Western Australia: The Government of Western Australia.

Babar, M. A., Ginkel, M. v., Klatt, A. R., Prasad, B., & Reynolds, M. P. (2006). The potential of using spectral reflectance indices to estimate yield in wheat grown under reduced irrigation. *Euphytica, 150*(1-2), 155-172. doi:https://doi.org/10.1007/s10681-006-9104-9

Barrett, R. L. (2016). *Perth plants: a field guide to the bushland and coastal flora of Kings Park and Bold Park* (2nd ed.). Clayton South, Victoria: CSIRO Publishing.

Basukala, A. K., Oldenburg, C., Schellberg, J., Sultanov, M., & Dubovyk, O. (2017). Towards improved land use mapping of irrigated croplands: performance assessment of different image classification algorithms and approaches. *European Journal of Remote Sensing, 50*(1), 187-201. doi:https://doi.org/10.1080/22797254.2017.1308235

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment, 204*, 509-523. doi:https://doi.org/10.1016/j.rse.2017.10.005

Belgiu, M., & Dra˘gut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing, 114*, 24-31. doi:http://doi.org/10.1016/j.isprsjprs.2016.01.011

Belward, A. S., & Skøien, J. O. (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing, 103*, 115-128. doi:https://doi.org/10.1016/j.isprsjprs.2014.03.009

Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., & Heynen, M. (2004). Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing, 58*(3-4), 239-258. doi:https://doi.org/10.1016/j.isprsjprs.2003.10.002

Bischof, H., Schneider, W., & Pinz, A. J. (1992). Multispectral classification of Landsat-images using neural networks. *IEEE, 30*(3), 482-490. doi:https://doi.org/10.1109/36.142926

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*(1), 2-16. doi:https://doi.org/10.1016/j.isprsjprs.2009.06.004

Bode, M., Brennan, K. E. C., Helmstedt, K., Desmond, A., Smia, R., & Algar, D. (2013). Interior fences can reduce cost and uncertainty when eradicating invasive species from large islands. *Methods in Ecology and Evolution, 4*(9), 819-827. doi:https://doi.org/10.1111/2041-210X.12072

Bolyn, C., Michez, A., Gaucher, P., Lejeune, P., & Bonnet, S. (2018). Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. *Biotechnologie, Agronomie, Société et Environnement, 22*(3), 172-187.

Booth, T. D., Cox, S. E., Fifield, C., Phillips, M., & Williamson, N. (2005). Image analysis compared with other methods for measuring ground cover. *Arid Land Research and Management, 19*(2), 91-100. doi:https://doi.org/10.1080/15324980590916486

Bowder, S. (1990). Before Dirk Hartog: prehistoric archaeoidgical research in Shark Bay, Western Australia. *Australian Archaeology, 30*, 46-57. doi:https://doi.org/10.1080/03122417.1990.11681366

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate Analysis of Variance*. Newbury Park, California: Sage Publications.

Breiman, L. (1984). *Classification and regression trees* (1st ed.). New York, New York: Routledge.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:https://doi.org/10.1023/A:1010933404324

Burbidge, A. A., & Woinarski, J. (2016a). Lagorchestes hirsutus. Retrieved from https://www.iucnredlist.org/species/11162/21954429

Burbidge, A. A., & Woinarski, J. (2016b). Lagostrophus fasciatus. Retrieved from https://www.iucnredlist.org/species/11171/21955969

Bureau of Meteorology. (2019). Climate Data Online. Retrieved from http://www.bom.gov.au/climate/data/

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2*(2), 121-167. doi:https://doi.org/10.1023/A:1009715923555

Burnett, C., & Blaschke, T. (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling, 168*, 233-249. doi:http://doi.org/10.1016/S0304-3800(03)00139-X

Campbell, S., Roberts, E. J., Craemer, R., Pacioni, C., Rollins, L., & Woolnough, A. P. (2016). Assessing the economic benefits of starling detection and control to Western Australia. *Australasian Journal of Environmental Management, 23*(1), 81-99. doi:https://doi.org/10.1080/14486563.2015.1028486

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., & Malo, J. (2011). *Remote sensing image processing*. LaPorte, Colorado: Morgan & Claypool.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances, 1*(5), 1-6. doi:https://doi.org/10.1126/sciadv.1400253

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *PNAS, 114*(30), E6089-E6096. doi:https://doi.org/10.1073/pnas.1704949114

Champion, I., Germain, C., Costa, J. P. D., Alborini, A., & Dubois-Fernandez, P. (2013). Retrieval of Forest Stand Age From SAR Image Texture for Varying Distance and Orientation Values of the Gray Level Co-Occurrence Matrix. *IEEE Geoscience and Remote Sensing Letters, 11*(1), 5-9. doi:https://doi.org/10.1109/LGRS.2013.2244060

Chapman, D. S., Bonn, A., Kunin, W. E., & Cornell, S. J. (2010). Random Forest characterization of upland vegetation and management burning from aerial imagery. *Journal of Biogeography, 37*(1), 37-46. doi:https://doi.org/10.1111/j.1365-2699.2009.02186.x

Chavez, P. S., & Bauer, B. (1982). An automatic optimum kernel-size selection technique for edge enhancement. *Remote Sensing of Environment, 12*(1), 23-28. doi:https://doi.org/10.1016/0034-4257(82)90005-0

Chen, Y., Luo, M., Xu, L., Zhou, X., Ren, J., & Zhou, J. (2018). *Object-based random forest classification of land cover from remotely sensed imagery for industrial and mining reclamation.* Paper presented at the ISPRS TC III Mid Term Symposium 'Developments, Technologies and Applications in Remote Sensing', Beijing, China.

Chianucci, F., Disperati, L., Guzzi, D., Bianchini, D., Nardino, V., Lastri, C., . . . Corona, P. (2016). Estimation of canopy attributes in beech forests using true colour digital images from a small fixed-wing UAV. *International Journal of Applied Earth Observation and Geoinformation, 47*, 60-68. doi:http://doi.org/10.1016/j.jag.2015.12.005

Clarke, G. S., Crossland, M. R., & Shine, R. (2016). Can we control the invasive cane toad using chemicals that have evolved under intraspecific competition? *Ecological Applications, 26*(2), 463-474. doi:https://doi.org/10.1890/14-2365

Clausi, D. A. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing, 28*(1), 45-62. doi:https://doi.org/10.5589/m02-004

Coburn, C. A., & Roberts, A. C. B. (2004). A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing, 25*(20), 4287-4308. doi:https://doi.org/10.1080/0143116042000192367

Coleman, H. (1998). Triodia plurinervata N.T.Burb. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/694

Colkesen, I., & Kavzoglu, T. (2018). Selection of optimal object features in object-based image analysis using filter-based algorithms. *Journal of the Indian Society of Remote Sensing, 46*(8), 1233-1242. doi:https://doi.org/10.1007/s12524-018-0807-x

Congalton, R. G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data: principles and practices* (2nd ed.). Boca Raton, Florida: CRC Press.

Corporation, M. (2013). Microsoft Excel (Version 15.0.5179.1000). Redmond, Washington: Microsoft Corporation.

Cowen, S., Rayner, K., Sims, C., & Morris, K. (2018). *Dirk Hartog Island National Park ecological restoration project: stage one - trial hare-wallaby translocations and monitoring*. Perth: Government of Western Australia.

Crooks, J. A. (2002). Characterizing ecosystem-level consequences of biological invasions: the role of ecosystem engineers. *Oikos, 97*(2), 153-166. doi:https://doi.org/10.1034/j.1600-0706.2002.970201.x

Cuevas, M. F., Mastrantonio, L., Ojeda, R. A., & Jaksic, F. M. (2012). Effects of wild boar disturbance on vegetation and soil properties in the Monte Desert, Argentina. *Mammalian Biology, 77*(4), 299-306. doi:https://doi.org/10.1016/j.mambio.2012.02.003

Cutler, D. R., Edwards, T. C., Beard, K. H., & Cutler, A. (2007). Random Forests for Classification in Ecology. *Ecology, 88*(11), 2783-2792. doi:https://doi.org/10.1890/07-0539.1

Czernecki, B., Nowosad, J., & Jabłońska, K. (2018). Machine learning modeling of plant phenology based on coupling satellite and gridded meteorological dataset. *International Journal of Biometeorology, 62*(7), 1297-1309. doi:https://doi.org/10.1007/s00484-018-1534-2

Davidson, A. D., Detling, J. K., & Brown, J. H. (2012). Ecological roles and conservation challenges of social, burrowing, herbivorous mammals in the world's grasslands. *Frontiers in Ecology and the Environment, 10*(9), 477-486. doi:https://doi.org/10.1890/110054

Davis, E. R. (2018). Texture analysis *Computer vision: principles, algorithms, applications, learning* (5th ed., pp. 185-200). London, England: Academic Press.

DBCA. (2017). Threatened species introduced to Dirk Hartog Island National Park. Retrieved from https://www.dpaw.wa.gov.au/news/media-statements/minister-for-environment/item/3319-threatened-species-introduced-to-dirk-hartog-island-national-park

Dekavalla, M., & Argialas, D. (2018). A region merging segmentation with local scale parameters: applications to spectral and elevation data. *Remote Sensing, 10*(2024), 1-20. doi:https://doi.org/10.3390/rs10122024

Díaz-Varela, R. A., Iglesias, S. C., Castro, C. C., & Varelad, E. R. D. (2018). Sub-metric analisis of vegetation structure in bog-heathland mosaics using very high resolution rpas imagery. *Ecological Indicators, 89*, 861-873. doi:https://doi.org/10.1016/j.ecolind.2017.11.068

Dragulescu, A. A., & Arendt, C. (2018). xlsx: read, write, format Excel 2007 and Excel 97/2000/XP/2003 files (Version 0.6.1): R package. Retrieved from https://github.com/dragua/rexcel

Drăguţ, L., Csillik, O., Eisank, C., & Tiede, D. (2014). Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing, 88*, 119-127. doi:https://doi.org/10.1016/j.isprsjprs.2013.11.018

Drăguţ, L., Tiede, D., & Levick, S. R. (2010). ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science, 24*(6), 859-871. doi:https://doi.org/10.1080/13658810903174803

Dronova, I. (2015). Object-based image analysis in wetland research: a review. *Remote Sensing, 7*(5), 6380-6413. doi:https://doi.org/10.3390/rs70506380

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. doi:https://doi.org/10.1016/j.isprsjprs.2015.03.002

Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). Multi-scale object-based image analysis. *International Journal of Remote Sensing, 33*(14), 4502-4526. doi:https://doi.org/10.1080/01431161.2011.649864

Ehrlich, P. R., & Ehrlich, A. H. (2013). Can a collapse of global civilization be avoided? *Proceedings: Biological Sciences, 280*(1754), 1-9. doi:http://doi.org/10.1098/rspb.2012.2845

Eisenhauer, N., Milcu, A., Sabais, A. C. W., & Scheu, S. (2008). Animal ecosystem engineers modulate the diversity-invasibility relationship. *PLoS One, 3*(10), 1-8. doi:https://doi.org/10.1371/journal.pone.0003489

El-naggar, A. M. (2018). Determination of optimum segmentation parameter values for extracting building from remote sensing images. *Alexandria Engineering Journal, 57*(4), 3089-3097. doi:https://doi.org/10.1016/j.aej.2018.10.001

Eldridge, D. J., & James, A. I. (2009). Soil-disturbance by native animals plays a critical role in maintaining healthy Australian landscapes. *Ecological Management & Restoration, 10*(s1), s27-s34. doi:https://doi.org/10.1111/j.1442-8903.2009.00452.x

Eldridge, D. J., & Mensinga, A. (2007). Foraging pits of the short-beaked echidna (Tachyglossus aculeatus) as small-scale patches in a semi-arid Australian box woodland. *Soil Biology and Biochemisty, 5*(39), 1055-1065. doi:https://doi.org/10.1016/j.soilbio.2006.11.016

Espindola, G. M., Câmara, G., Reis, I. A., Leonardo, B., & Monteiro, A. M. V. (2006). Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *International Journal of Remote Sensing, 27*(14), 3035-3040. doi:https://doi.org/10.1080/01431160600617194

ESRI. (2019). ArcGIS Pro. Redlands, California: Environmental Systems Research Institute.

Falster, D. S., & Westoby, M. (2003). Plant height and evolutionary games. *Trends in Ecology and Evolution, 18*(7), 337-343. doi:http://doi.org/10.1016/S0169-5347(03)00061-2

Farrag, A. E.-H. A., Megahed, H. A., & Darwish, M. H. (2019). Remote sensing, GIS and chemical analysis for assessment of environmental impacts on rising of groundwater around Kima Company, Aswan, Egypt. *Bulletin of the National Research Centre, 43*(14). doi:https://doi.org/10.1186/s42269-019-0056-3

Feng, Q., Liu, J., & Gong, J. (2015). UAV remote sensing for urban vegetation mapping using Random Forest and texture analysis. *Remote Sensing, 7*(1), 1074-1094. doi:https://doi.org/10.3390/rs70101074

Fern, R. R., Foxley, E. A., Bruno, A., & Morrison, M. L. (2018). Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland. *Ecological Indicators, 94*(1), 16-21. doi:https://doi.org/10.1016/j.ecolind.2018.06.029

Fleming, P. A., Anderson, H., Prendergast, A. S., Bretz, M. R., Valentine, L. E., & Hardy, G. E. S. J. (2014). Is the loss of Australian digging mammals contributing to a deterioration in ecosystem function? *Mammal Review, 44*(2), 94-108. doi:https://doi.org/10.1111/mam.12014

Foody, G. M. (2004). Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing, 70*(5), 627-633. doi:https://doi.org/10.14358/PERS.70.5.627

Fugara, A. M. A., Pradhan, B., & Mohamed, T. A. (2009). Improvement of land-use classification using object-oriented and fuzzy logic approach. *Applied Geomatics, 1*, 111. doi:https://doi.org/10.1007/s12518-009-0011-3

Fujiwara, M., & Takada, T. (2017). Environmental Stochasticity *eLS*. Chichester, England: John Wiley & Son.

Garnett, S., Latch, P., Lindenmayer, D., & Woinarksi, J. (2018). *Recovering Australian threatened species* (1st ed.). Clayton South, Victoria: CSIRO.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3-42. doi:https://doi.org/10.1007/s10994-006-6226-1

Ghosh, A., Fassnacht, F. E., Joshi, P. K., & Koch, B. (2014). A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across

three spatial scales. *International Journal of Applied Earth Observation and Geoinformation, 26*, 49-63. doi:https://doi.org/10.1016/j.jag.2013.05.017

Gleason, A. C. R., Shihavuddin, A., Gracias, N., Schultz, G., & Gintert, B. E. (2015). *Improved supervised classification of underwater military munitions using height features derived from optical imagery.* Paper presented at the OCEANS 2015 - MTS/IEEE, Washington, District of Columbia.

Grossman, B. F., Hayward, M. W., & Gibb, H. (2019). An experimental test of multi-scalar impacts of digging mammal reintroductions on invertebrate burrows. *Soil Biology and Biochemistry, 132*, 101-110. doi:https://doi.org/10.1016/j.soilbio.2019.02.003

Guern, P. L., & Davaud, E. (2005). Recognition of ancient carbonate wind deposits: lessons from a modern analogue, Chrissi Island, Crete. *Sedimentology, 52*(5), 915-926. doi:https://doi.org/10.1111/j.1365-3091.2005.00700.x

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*(7-8), 1157-1182. doi:https://doi.org/10.1162/153244303322753616

Haining, R. P. (2015). Spatial Sampling. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed.): Elsevier Science Ltd.

Haralick, R. M., Shanmugam, K., & Dinstein, I. h. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-3*(6), 610-621. doi:https://doi.org/10.1109/TSMC.1973.4309314

Hardman, B., & Moro, D. (2006). Importance of diurnal refugia to a hare-wallaby reintroduction in Western Australia. *Wildlife Research, 33*(5), 355–359. doi:http://doi.org/10.1071/WR05088

Harvey, A., Johnson, M. E., & Harvey, R. (2018). Heterozoan carbonate-enriched beach sand and coastal dunes—with particular reference to rhodoliths, Dirk Hartog Island, Shark Bay, Western Australia. *Facies, 64*(3), 1-18. doi:https://doi.org/10.1007/s10347-018-0533-4

Hedin, L. O., Mayer, M. S., & Likens, G. E. (1988). The effect of deforestation on organic debris dams. *Internation Association for Theoretical and Applied Limnology, 23*(2), 1135-1141. doi:https://doi.org/10.1080/03680770.1987.11899783

Heriot, S., Asher, J., Williams, M. R., & Moro, D. (2019). The eradication of ungulates (sheep and goats) from DirkHartog Island, Shark Bay World Heritage Area, Australia. *Biological Invasions, 21*(5), 1789-1805. doi:https://doi.org/10.1007/s10530-019-01937-7

Hijmans, R. J. (2019). raster: geographic data analysis and modeling (Version 3.0-7): R package. Retrieved from https://CRAN.R-project.org/package=raster

Hilmer, S., Algar, D., & Johnston, M. (2010). Opportunistic observation of predation of Loggerhead turtle hatchlings by feral cats on Dirk Hartog Island, Western Australia. *Journal of the Royal Society of Western Australia, 93*(3), 141-146.

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE, 20*(8), 832-844. doi:https://doi.org/10.1109/34.709601

Hobbs, R. J. (1993). Effects of landscape fragmentation on ecosystem processes in the Western Australian wheatbelt. *Biological Conservation, 64*(3), 193-201. doi:https://doi.org/10.1016/0006-3207(93)90321-Q

Hobbs, R. J., & Hopkins, A. J. M. (1990). From frontier to fragments: European impact on Australia's vegetation. *Proceedings of the Ecological Society of Australia, 15*, 93-114.

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., . . . Zhou, G. (2018). Impacts of 1.5ºC Global Warming on Natural and Human Systems. In V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, & T. Waterfield (Eds.), *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global*

greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Geneva, Switzerland: World Meteorological Organization.

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., . . . O'Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature, 486*, 105-109. doi:https://doi.org/10.1038/nature11118

Huang, X., & Zhang, L. (2008). An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing, 46*(12), 4173-4185. doi:https://doi.org/10.1109/TGRS.2008.2002577

Huang, Y., Liu, X., Li, X., Yan, Y., & Ou, J. (2018). Comparing the Effects of Temporal Features Derived From Synthetic Time-Series NDVI on Fine Land Cover Classification. *IEEE journal of selected topics in applied earth observations and remote sensing, 11*(12), 4618-4629. doi:http://doi.org/10.1109/JSTARS.2018.2869528

Iannella, A., Peacock, D., Cassey, P., & Schwensow, N. (2019). Genetic perspectives on the historical introduction of the European rabbit (Oryctolagus cuniculus) to Australia. *Biological Invasions, 21*(2), 603-614. doi:https://doi.org/10.1007/s10530-018-1849-2

James, A. I., Eldridge, D. J., & Hill, B. M. (2009). Foraging animals create fertile patches in an Australian desert shrubland. *Ecography, 32*(5), 723-732. doi:https://doi.org/10.1111/j.1600-0587.2009.05450.x

Jebur, M. N., Shafri, H. Z. M., Pradhan, B., & Tehrany, M. S. (2014). Per-pixel and object-oriented classification methods for mapping urban land cover extraction using SPOT 5 imagery. *Geocarto International, 29*(7), 792-806. doi:https://doi.org/10.1080/10106049.2013.848944

Jhonnerie, R., Siregar, V. P., Nababan, B., Prasetyo, L. B., & Wouthuyzen, S. (2015). Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and Alos Palsar Imageries. *Procedia Environmental Sciences, 24*, 215-221. doi:https://doi.org/10.1016/j.proenv.2015.03.028

JianYang, Li, P., & He, Y. (2014). A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing, 94*, 13-24. doi:https://doi.org/10.1016/j.isprsjprs.2014.04.008

Jiao, L., & Liu, Y. (2012). *Analyzing the shape characteristics of land use classes in remote sensing imagery*. Paper presented at the XXII ISPRS Congress, Melbourne, Victoria.

Jin, C., & Liu, J.-A. (2010). *Applications of Support Vector Machine and unsupervised learning for predicting maintainability using object-orientated metrics*. Paper presented at the Second International Conference of MultiMedia and Information Technology, Kaifeng, China.

Jin, Y., Liu, X., Chen, Y., & Liang, X. (2018). Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. *International Journal of Remote Sensing, 39*(23), 8703-8723. doi:https://doi.org/10.1080/01431161.2018.1490976

Jing, L., Hu, B., Noland, T., & Li, J. (2012). An individual tree crown delineation method based on multi-scale segmentation of imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 70*, 88-98. doi:https://doi.org/10.1016/j.isprsjprs.2012.04.003

Johnson, B., & Xie, Z. (2011). Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(4), 473-483. doi:https://doi.org/10.1016/j.isprsjprs.2011.02.006

Johnson, B. A., Bragais, M., Endo, I., Magcale-Macandog, D. B., & Macandog, P. B. M. (2015). Image segmentation parameter optimization considering within- and between-segment heterogeneity at multiple scale levels: test case for mapping residential areas using landsat

imagery. *ISPRS International Journal of Geo-Information, 4*(4), 2292-2305. doi:https://doi.org/10.3390/ijgi4042292

Jones, C. G., Gutiérrez, J. L., Byers, J. E., Crooks, J. A., Lambrinos, J. G., & Talley, T. S. (2010). A framework for understanding physical ecosystem engineering by organisms. *Oikos, 119*(12), 1862-1869. doi:https://doi.org/10.1111/j.1600-0706.2010.18782.x

Jones, C. G., Lawton, J. H., & Shachak, M. (1994). Organisms as ecosystem engineers. *Oikos, 69*(3), 373-386. doi:https://doi.org/10.2307/3545850

Jones, C. G., Lawton, J. H., & Shachak, M. (1997). Ecosystem engineering by organisms: why semantics matters. *Trends in Ecology and Evolution, 12*(7), 275. doi:https://doi.org/10.1016/S0169-5347(97)81019-1

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-269. doi:https://doi.org/10.1126/science.aaa8415

Juniati, E., & Arrofiqoh, E. N. (2017). *Comparison of pixel-based and object-based classification using parameters and non-parameters approach for the pattern consistency of multi scale landcover*. Paper presented at the The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Wuhan, China.

Kavzoglu, T., & Mather, P. (2002). The role of feature selection in artficial neural network applications. *International Journal of Remote Sensing, 23*(15), 2919-2937. doi:https://doi.org/10.1080/01431160110107743

Kavzoglu, T., & Yildiz, M. E. (2014). *Parameter-based performance analysis of object-based image analysis using aerial and Quickbird-2 images*. Paper presented at the ISPRS Technical Commission VII Symposium, Istanbul, Turkey.

Kawakubo, F. S., Morato, R. G., & Luchiari, A. (2013). Use of fraction imagery, segmentation and masking techniques to classify land-use and land-cover types in the Brazilian Amazon. *International Journal of Remote Sensing, 34*(15), 5452-5467. doi:https://doi.org/10.1080/01431161.2013.791758

Kelly, M., Blanchard, S. D., Kersten, E., & Koy, K. (2011). Terrestrial remotely sensed imagery in support of public health: new avenues of research using object-based image analysis. *Remote Sensing, 3*(11), 2321-2345. doi:https://doi.org/10.3390/rs3112321

Kim, M., Madden, M., & Warner, T. A. (2008). Estimation of optimal image object size for the segmentation of forest stands with multispectral IKONOS imagery. In T. Blaschke, S. Lang, & G. Hay (Eds.), *Object-Based Image Analysis. Lecture Notes in Geoinformation and Cartography*. Berlin, Germany: Springer.

Kim, M., Madden, M., & Warner, T. A. (2009). Forest type mapping using object-specific texture measures from multispectral Ikonos imagery: segmentation quality and image classification issues. *Photogrammetric Engineering and Remote Sensing, 75*(7), 819-829. doi:https://doi.org/10.14358/PERS.75.7.819

Kim, M., Warner, T. A., Madden, M., & Atkinson, D. S. (2011). Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *International Journal of Remote Sensing, 32*(10), 2825-2850. doi:https://doi.org/10.1080/01431161003745608

Klosterman, S., Melaas, E., Wang, J. A., Martinez, A., Frederick, S., O'Keefe, J., . . . Richardson, A. D. (2018). Fine-scale perspectives on landscape phenology from unmanned aerial vehicle (UAV) photography. *Agricultural and Forest Meteorology, 248*, 397-407. doi:http://doi.org/10.1016/j.agrformet.2017.10.015

Kluckner, S., Mauthner, T., Roth, P. M., & Bischof, H. (2009). *Semantic classification in aerial imagery by integrating appearance and height information*. Paper presented at the 9th Asian Conference on Computer Vision, Xi'an, China.

Kuhn, M. (2019). caret: classification and regression training (Version 6.0-84): R package. Retrieved from https://github.com/topepo/caret/

Kumar, P., Prasad, R., Choudhary, A., Mishra, V. N., Gupta, D. K., & Srivastava, P. K. (2017). A statistical significance of differences in classification accuracy of crop types using different classification algorithms. *Geocarto International, 32*(2), 206-224. doi:http://doi.org/10.1080/10106049.2015.1132483

Kumar, U., Dasgupta, A., Mukhopadhyay, C., & Ramachandra, T. V. (2018). Examining the Effect of Ancillary and Derived Geographical Data on Improvement of Per-Pixel Classification Accuracy of Different Landscapes. *Journal of the Indian Society of Remote Sensing, 46*(3), 407-422. doi:https://doi.org/10.1007/s12524-017-0698-2

Levin, S. A. (2009). *The Princeton Guide to Ecology*. Princeton, New Jersey: Princeton University Press.

Lillesand, T. M., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation*. New York, New York: John Wiley & Sons.

Lindenmayer, D. (2007). *On borrowed time: Australia's environmental crisis and what we must do about it*. Camberwell, Victoria: Penguin Group.

Linder, P. H., Bykova, O., Dyke, J., Etienne, R. S., Hickler, T., Kühn, I., . . . Singer, A. (2012). Biotic modifiers, environmental modulation and species distribution models. *Journal of Biogeography, 39*(12). doi:https://doi.org/10.1111/j.1365-2699.2012.02705.x

Liu, E.-q., Zhou, W.-c., Zhou, J.-m., Shao, H.-y., & Yang, X. (2013). Combining spectral with texture features into object oriented classification in mountainous terrain using advanced land observing satellite image. *Journal of Mountain Science, 10*(5), 768-776. doi:https://doi.org/10.1007/s11629-013-2532-9

Liu, X., Xu, J., Zhao, J., Yong, L., & Xin, Z. (2015). The comparison of segmentation results for high-resolution remote sensing image between eCognition and EDISON. *Applied Mechanics and Materials, 713-715*, 373-376. doi:https://doi.org/10.4028/www.scientific.net/AMM.713-715.373

Liu, X., & Yetik, I. S. (2010). *A Maximum Likelihood Classification method for image segmentation considering subject variability*. Paper presented at the Southwest Symposium on Image Analysis & Interpretation, Austin, Texas.

Liua, K., Shia, W., & Zhang, H. (2011). A fuzzy topology-based maximum likelihood classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(1), 103-114. doi:https://doi.org/10.1016/j.isprsjprs.2010.09.007

Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto International, 16*(1), 65-70. doi:https://doi.org/10.1080/10106040108542184

Louw, M. A., Roux, P. C. l., Meyer-Milne, E., & Haussmann, N. S. (2017). Mammal burrowing in discrete landscape patches further increases soil and vegetation heterogeneity in an arid environment. *Journal of Arid Environments, 141*. doi:https://doi.org/10.1016/j.jaridenv.2017.02.007

Lu, D., Hetrick, S., & Moran, E. (2010). Land cover classification in a complex urban-rural landscape with QuickBird imagery. *Photogrammetric Engineering and Remote Sensing, 76*(10), 1159-1168. doi:https://doi.org/10.14358/PERS.76.10.1159

Lu, D., Hetrick, S., & Moran, E. (2011). Impervious surface mapping with Quickbird imagery. *International Journal of Remote Sensing, 32*(9), 2519-2533. doi:https:/doi.org/10.1080/01431161003698393

Lu, D., & Weng, Q. (2007). Survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing, 28*(5), 823-870. doi:https://doi.org/10.1080/01431160600746456

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 130*, 277-293. doi:https://doi.org/10.1016/j.isprsjprs.2017.06.001

Macfarlane, C., & Ogden, G. N. (2012). Automated estimation of foliage cover in forest understorey from digital nadir images. *Methods in Ecology and Evolution, 3*(2). doi:https://doi.org/10.1111/j.2041-210X.2011.00151.x

Magnussen, S., McRoberts, R. E., & Tomppo, E. (2009). Model-based mean square error estimators for k-nearest neighbor predictions and applications using remotely sensed data for forest inventories. *Remote Sensing of Environment, 113*(3), 476-488. doi:http://doi.org/10.1016/j.rse.2008.04.018

Malahlela, O., Cho, M. A., & Mutanga, O. (2014). Mapping canopy gaps in an indigenous subtropical coastal forest using high-resolution WorldView-2 data. *International Journal of Remote Sensing, 35*(17), 6397-6417. doi:https://doi.org/10.1080/01431161.2014.954061

Mallen-Cooper, M., Nakagawa, S., & Eldridge, D. J. (2019). Global meta-analysis of soil-disturbing vertebrates reveals strong effects on ecosystem patterns and processes. *Global Ecology and Biogeography, 28*(5), 661-679. doi:https://doi.org/10.1111/geb.12877

Man, Q., Dong, P., & Guo, H. (2015). Pixel- and feature-level fusion of hyperspectral and lidar data for urban land-use classification. *International Journal of Remote Sensing, 36*(6), 1618-1644. doi:https://doi.org/10.1080/01431161.2015.1015657

Manning, A. D., Eldridge, D. J., & Jones, C. G. (2015). Policy implications of ecosystem engineering for multiple ecosystem benefits. In D. P. Armstrong, M. W. Hayward, D. Moro, & P. J. Seddon (Eds.), *Advances in reintroduction biology of Australian and New Zealand Fauna* (pp. 167-184). Clayton South, Victoria: CSIRO Publishing.

Maryan, B. (1996). Herpetofauna of Dirk Hartog Island Shark Bay area, Western Australia. *Herpetofauna, 26*(1), 8-11.

Mathieu, R., Aryal, J., & Chong, A. K. (2007). Object-based classification of Ikonos imagery for mapping large-scale vegetation communities in urban areas. *Sensors, 7*(11), 2860-2880. doi:https://doi.org/10.3390/s7112860

Matthew L Brooks, D'Antonio, C. M., Richardson, D. M., Grace, J. B., Keeley, J. E., DiTomaso, J. M., . . . Pyke, D. (2004). Effects of invasive alien plants on fire regimes. *BioScience, 54*(7), 677-688. doi:https://doi.org/10.1641/0006-3568(2004)054[0677:EOIAPO]2.0.CO;2

McInerney, D., Kempeneers, P., Marron, M., & McRoberts, R. E. (2019). Analysis of broadleaf encroachment in coniferous forest plantations using multi-temporal satellite imagery. *International Journal of Applied Earth Observation and Geoinformation, 78*, 130-137. doi:https://doi.org/10.1016/j.jag.2018.12.005

Melesse, A. M., Weng, Q., Thenkabail, P. S., & Senay, G. B. (2007). Remote sensing sensors and applications in environmental resources mapping and modelling. *Sensors, 7*(12), 3209-3241. doi:https://doi.org/10.3390/s7123209

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

Memarian, H., Balasundram, S. K., & Khosla, R. (2013). Comparison between pixel- and object-based image classification of a tropical landscape using Système Pour l'Observation de la Terre-5 imagery. *Journal of Applied Remote Sensing, 7*(1), 073512. doi:https://doi.org/10.1117/1.JRS.7.073512

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*(7), 8489-8515. doi:https://doi.org/10.3390/rs70708489

Mishra, V. N., Prasad, R., Kumar, P., Gupta, D. K., & Srivastava, P. K. (2017). Dual-polarimetric C-band SAR data for land use/land cover classification by incorporating textural information. *Environmental Earth Sciences, 76*(1), 1-16. doi:https://doi.org/10.1007/s12665-016-6341-7

Mitchell, T. M. (1997). *Machine Learning*. New York, New York: McGraw-Hill.

Möllera, M., Lymburner, L., & M.Volkc. (2007). The comparison index: A tool for assessing the accuracy of image segmentation. *International Journal of Applied Earth Observation and Geoinformation, 9*(3), 311-321. doi:https://doi.org/10.1016/j.jag.2006.10.002

Molles, M. C. (2009). *Ecology Concepts and Applications* (5th ed.). New York, New York: McGraw-Hill Education.

Mostafa, H., & Wang, X. (2019). *Parameter efficient training of deep convolutional neural networks by dynamic sparse reparameterization*. Paper presented at the International Conference on Machine Learning, Long Beach, California.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 66*(3), 247-259. doi:https://doi.org/10.1016/j.isprsjprs.2010.11.001

Müllerová, J., Bartaloš, T., Brůna, J., Dvořák, P., & Vítková, M. (2017). Unmanned aircraft in nature conservation: an example from plant invasions. *International Journal of Remote Sensing, 38*(8-10), 2177-2198. doi:https://doi.org/10.1080/01431161.2016.1275059

Mulligan, M., Buxton, M., Lane, R., Neave, M., & Richardson, A. (2015). *An introduction to sustainability: environmental, social and personal* New York, New York: Routledge.

Munro, N. T., McIntyre, S., Macdonald, B., Cunningham, S. A., Gordon, I. J., Cunningham, R. B., & Manning, A. D. (2019). Returning a lost process by reintroducing a locally extinct digging marsupial. *PeerJ*. doi:https://doi.org/10.7717/peerj.6622

Myint, S. W., Gober, P., Brazel, A., Grossman-Clarke, S., & Weng, Q. (2011). Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment, 115*(5), 1145-1161. doi:https://doi.org/10.1016/j.rse.2010.12.017

Nettleton, D. (2014). Selection of Variables and Factor Derivation *Commercial Data Mining: processing, analysis and modeling for predictive analytics projects* (pp. 79-104): Elsevier Science Ltd.

Niphadkar, M., Nagendra, H., Tarantino, C., Adamo, M., & Blonda, P. (2017). Comparing Pixel and Object-Based Approaches to Map an Understorey Invasive Shrub in Tropical Mixed Forests. *Frontiers in Plant Science, 8*(892). doi:https://doi.org/10.3389/fpls.2017.00892

Nogueira, K., Santos, J. A. d., Menini, N., Silva, T. S. F., Morellato, L. P. C., & Torres, R. d. S. (2019). Spatio-temporal vegetation pixel classification by using convolutional networks. *IEEE Geoscience and Remote Sensing Letters, 1*(1), 1-5. doi:https:/doi.org/10.1109/LGRS.2019.2903194

Noi, P. T., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors, 18*(1), 18. doi:https://doi.org/10.3390%2Fs18010018

Ochoa, K. S., & Guo, Z. (2019). A framework for the management of agricultural resources with automated aerial imagery detection. *Computers and Electronics in Agriculture, 162*, 53-69. doi:https://doi.org/10.1016/j.compag.2019.03.028

Olsen, R. C. (2007). EO - Spectral Imagery *Remote Sensing from Air and Space* (pp. 117-136). Washington, District of Columbia: Society of Photo-optical Instrumentation Engineers.

Paczkowska, G. (1993). Cenchrus ciliaris L. Buffel Grass. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/258

Paczkowska, G. (1994). Acanthocarpus preissii Lehm. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/1208

Paczkowska, G. (1995a). Exocarpos aphyllus R.Br. Leafless Ballart. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/10977

Paczkowska, G. (1995b). Leptospermum laevigatum (Gaertn.) F.Muell. Coast Teatree. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/5850

Paczkowska, G. (1996). Alyogyne cuneiformis (DC.) Lewton Coastal Hibiscus. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/4904

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing, 26*(1), 217-222. doi:https://doi.org/10.1080/01431160412331269698

Pantaleoni, E., Wynne, R., Galbraith, J., & Campbell, J. B. (2009). Mapping wetlands using ASTER data: A comparison between classification trees and logistic regression. *International Journal of Remote Sensing, 30*(13). doi:https:/doi.org/10.1080/01431160802562214

Pavlidis, G. (2017). *Mixed raster content: segmentation, compression, transmission* (1st ed.). Singapore, Republic of Singapore: Springer.

Pebesma, E. (2019). sf: simple features for R (Version 0.8-0): R package. Retrieved from https://r-spatial.github.io/sf/

Pelt, D. M., & Sethian, J. A. (2018). A mixed-scale dense convolutional neural network for image analysis. *PNAS, 115*(2), 254-259. doi:https://doi.org/10.1073/pnas.1715832114

Peterson, B. G. (2019). PerformanceAnalytics: econometric tools for performance and risk analysis (Version 1.5.3): R package. Retrieved from https://github.com/braverock/PerformanceAnalytics

Phillips, J. D. (2013). Human impacts on the environment: unpredicatbility and the primacy of place. *Physical Geography, 22*(4), 321-332. doi:https://doi.org/10.1080/02723646.2001.10642746

Pillai, K. C. S. (1955). Some new test criteria in Multivariate Analysis. *The Annals of Mathematical Statistics, 26*(1), 117-121.

Pittock, J., Finlayson, C. M., & Howitt, J. (2013). Beguiling and risky: 'environmental works and measures' for wetland conservation under a changing climate. *Hydrobiologia, 708*(1), 111-131. doi:https://doi.org/10.1007/s10750-012-1292-9

Pollock, M. M., Naiman, R. J., Erickson, H. E., Johnston, C. A., Pastor, J., & Pinay, G. (1995). Beaver as engineers: influences on biotic and abiotic characteristics of drainage basins. In C. G. Jones & J. H. Lawton (Eds.), *Linking Species & Ecosystems* (pp. 117-126). New York, New York: Springer.

Pu, R., & Landry, S. (2012). A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment, 124*, 516-533. doi:https://doi.org/10.1016/j.rse.2012.06.011

Puissant, A., Rougier, S., & Stumpf, A. (2014). Object-orientated mapping of urban trees using Random Forest. *International Journal of Applied Earth Observation and Geoinformation, 26*, 235-245. doi:https://doi.org/10.1016/j.jag.2013.07.002

Queensland Government. (2016). Environmental Weeds of Australia. Retrieved from http://keyserver.lucidcentral.org/weeds/data/03030800-0b07-490a-8d04-0605030c0f01/media/Html/Leptospermum_laevigatum.htm

R Studio. (2019). R Studio (Version 1.2.1335). Boston, Massachusetts: R Studio, Inc.

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX

images. *European Journal of Remote Sensing, 50*(1), 144-154. doi:https://doi.org/10.1080/22797254.2017.1299557

Radoux, J., & Bogaert, P. (2017). Good practices for object-based accuracy assessment. *Remote Sensing, 9*(646), 1-23. doi:https://doi.org/10.3390/rs9070646

Radoux, J., Bogaert, P., Fasbender, D., & Defourny, P. (2011). Thematic accuracy assessment of geographic object-based image classification. *International Journal of Geographical Information Science, 25*(6), 895-911. doi:https://doi.org/10.1080/13658816.2010.498378

Räsänen, A., Kuitunen, M., Tomppo, E., & Lensu, A. (2014). Coupling high-resolution satellite imagery with ALS-based canopy height model and digital elevation model in object-based boreal forest habitat type classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 94*, 169-182. doi:http://doi.org/10.1016/j.isprsjprs.2014.05.003

Redowan, M., Akter, S., & Islam, N. (2014). Analysis of forest cover change at Khadimnagar National Park, Sylhet, Bangladesh, using Landsat TM and GIS data. *Journal of Forestry Research, 25*(2), 393-400. doi:http://doi.org/10.1007/s11676-014-0467-9

Reubens, B., Poesen, J., Danjon, F., Geudens, G., & Muys, B. (2007). The role of fine and coarse roots in shallow slope stability and soil erosion control with a focus on root system architecture: a review. *Trees, 21*(4), 385-402. doi:http://doi.org/10.1007/s00468-007-0132-4

Richards, J. A. (2013). *Remote sensing digital image analysis: an introduction* (5th Ed.). Berlin, Heidelberg: Springer.

Richards, J. D. (2012). *Rufous hare-wallaby (Lagorchestes hirsutus) national recovery plan*. Bentley, Western Australia: Department of Environment and Conservation Retrieved from https://www.environment.gov.au/system/files/resources/e80dd3ce-9b10-4896-a351-10f1f004487e/files/rufous-hare-wallaby.pdf.

Robinson, T. P., Wardell-Johnson, G. W., Pracilio, G., Brown, C., Corner, R., & Klinken, R. D. v. (2016). Testing the discrimination and detection limits of WorldView-2 imagery on a challenging invasive plant target. *International Journal of Applied Earth Observation and Geoinformation, 44*, 23-30. doi:https://doi.org/10.1016/j.jag.2015.07.004

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment, 121*, 93-107. doi:https://doi.org/10.1016/j.rse.2011.12.003

Ross, C. E., Munro, N. T., Barton, P. S., Evans, M. J., Gillen, J., Macdonald, B. C. T., . . . Manning, A. D. (2019). Effects of digging by a native and introduced ecosystem engineer on soil physical and chemical properties in temperate grassy woodland. *PeerJ, 7*, e7506. doi:https://doi.org/10.7717/peerj.7506

Russo, A. G., Eden, J.-S., Tuipulotu, D. E., Shi, M., Selechnik, D., Shine, R., . . . White, P. A. (2018). Viral discovery in the invasive Australian cane toad (Rhinella marina) using metatranscriptomic and genomic approaches. *Journal of Virology, 92*(17), 1-18. doi:https://doi.org/10.1128/JVI.00768-18

Sabanci, K., Kayabasi, A., & Toktas, A. (2017). Computer vision-based method for classification of wheat grains using artificial neural network. *Journal of the Science of Food and Agriculture, 97*(8), 2588-2593. doi:https://doi.org/10.1002/jsfa.8080

Saboori, M., Torahi, A. A., & Bakhtayari, H. R. R. (2019). Combining multi-scale textural features from the panchromatic bands of high spatial resolution images with ANN and MLC classification algorithms to extract urban land uses. *International Journal of Remote Sensing, 40*(22), 8608-8634. doi:https://doi.org/10.1080/01431161.2019.1620371

Salehi, B., Zhang, Y., & Zhong, M. (2013). A combined object- and pixel-based image analysis framework for urban land cover classifiation of VHR imagery. *Photogrammetric Engineering and Remote Sensing, 79*(11), 999-1014. doi:https://doi.org/10.14358/PERS.79.11.999

Salehi, B., Zhang, Y., Zhong, M., & Dey, V. (2012). Object-based classification of urban areas using VHR

imagery and height points ancillary data. *Remote Sensing, 4*(8), 2256-2276. doi:http://doi.org/10.3390/rs4082256

Salembier, P., & Garrido, L. (1998). *Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval.* Paper presented at the Proceedings 1998 International Conference on Image Processing, Chicago, Illinois.

Sandom, C. J., Hughes, J., & Macdonald, D. W. (2013). Rooting for rewilding: quantifying wild boar's Sus scrofa rooting rate in the Scottish Highlands. *Restoration Ecology, 21*(3). doi:https://doi.org/10.1111/j.1526-100X.2012.00904.x

Sertel, E., & Alganci, U. (2015). Comparison of pixel and object-based classification for burned area mapping using SPOT-6 images. *Geomatics, Natural Hazards and Risk, 7*(4), 1198-1206. doi:https://doi.org/10.1080/19475705.2015.1050608

Sesnie, S. E., Finegan, B., Gessler, P. E., Thessler, S., Bendana, Z. R., & Smith, A. M. S. (2010). The multispectral separability of Costa Rican rainforest types with support vector machines and Random Forest decision trees. *International Journal of Remote Sensing, 31*(11), 2885-2909. doi:https://doi.org/10.1080/01431160903140803

Shahbahrami, A., Pham, T. A., & Bertels, K. (2012). Parallel implementation of Gray Level Co-occurrence Matrices and Haralick texture features on cell architecture. *The Journal of Supercomputing, 59*(3), 1455-1477. doi:https://doi.org/10.1007/s11227-011-0556-x

Shao, F., Long, F., Liang, J., Chen, H., & Yuan, M. (2015). A comparison of segmentation programs for high resolution remote sensing data. *Applied Mechanics and Materials, 713-715*, 373-376. doi:https://doi.org/10.4028/www.scientific.net/AMM.713-715.373

Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing, 70*, 78-87. doi:https://doi.org/10.1016/j.isprsjprs.2012.04.001

Sherba, J., Blesius, L., & Davis, J. D. (2014). Object-based classification of abandoned logging roads under heavy canopy using LiDAR. *Remote Sensing, 6*, 4043-4060. doi:https://doi.org/10.3390/rs6054043

Sibaruddin, H. I., Shafri, H., Pradhan, B., & Haron, N. A. (2018). Comparison of pixel-based and object-based image classification techniques in extracting information from UAV imagery data. *IOP Conference Series Earth and Environmental Science, 169*(1), 012098. doi:https://doi.org/10.1088/1755-1315/169/1/012098

Silveira, E. M. O., Silv, S. H. G., Acerbi-Junior, F. W., Carvalho, M. C., Carvalho, L. M. T., Scolforo, J. R. S., & Wulder, M. A. (2019). Object-based random forest modelling of aboveground forest biomass outperforms a pixel-based approach in a heterogeneous and mountain tropical environment. *International Journal of Applied Earth Observation and Geoinformation, 78*, 175-188. doi:https://doi.org/10.1016/j.jag.2019.02.004

Singh, M., Singh, S., & Partridge, D. (2005). Parameter optimization for image segmentation algorithms: a systematic approach. In S. Singh, M. Singh, C. Apte, & P. Perner (Eds.), *Pattern Recognition and Image Analysis: Third International Conference on Advances in Pattern Recognition* (Vol. 3687). Berlin, Germany: Springer.

Smith, J. H., Stehman, S. V., Wickham, J. D., & Yang, L. (2003). Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment, 84*(3), 342-349. doi:https://doi.org/10.1016/S0034-4257(02)00126-8

Spooner, A. (1999a). Acacia ligulata Benth. Umbrella Bush. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/3419

Spooner, A. (1999b). Atriplex vesicaria Benth. Bladder Saltbush. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/2481

Spooner, A. (2000). Scaevola spinescens R.Br. Currant Bush. Retrieved from https://florabase.dpaw.wa.gov.au/browse/profile/7644

Srivastava, N., Krizhevsky, G. H. A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*(1), 1929-1958.

Streitberger, M., & Fartmann, T. (2016). Vegetation heterogeneity caused by an ecosystem engineer drives oviposition-site selection of a threatened grassland insect. *Arthropod-Plant Interactions, 10*, 545-555. doi:http://doi.org/10.1007/s11829-016-9460-x

Tabb, M., & Ahuja, N. (1997). Multiscale image segmentation by integrated edge and region detection. *IEEE Transactions on Image Processing, 6*(5), 642-655. doi:https://doi.org/10.1109/83.568922

Talavera, L. (2005). *An evaluation of filter and wrapper methods for feature selection in categorical clustering.* Paper presented at the Advances in Intelligent Data Analysis VI. IDA, Berlin, Germany.

Tang, L., & Shao, G. (2015). Drone remote sensing for forestry research and practices. *Journal of Forestry Research, 26*(4), 791-797. doi:https://doi.org/10.1007/s11676-015-0088-y

Taylor, B. D., & Goldingay, R. L. (2010). Roads and wildlife: impacts, mitigation, and implications for wildlife management in Australia. *Wildlife Research, 37*(4), 320-331. doi:https://doi.org/10.1071/WR09171

Tehrany, M. S., Pradhan, B., & Jebuv, M. N. (2013). A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using SPOT 5 imagery. *Geocarto International, 29*(4), 351-369. doi:https://doi.org/10.1080/10106049.2013.768300

Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., . . . Huete, A. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing, 144*, 325-340. doi:https://doi.org/10.1016/j.isprsjprs.2018.07.017

The R Foundation. (2019). R (Version 3.6.1). Vienna, Austria: The R Foundation for Statistical Computing.

Thessen, A. E. (2016). Adoption of Machine Learning Techniques in Ecology and Earth Science. *One Ecosystem, 1*, 1-38. doi:https://doi.org/10.3897/oneeco.1.e8621

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography, 46*, 234-240. doi:http://doi.org/10.2307/143141

Townshend, J. R., Huang, C., Kalluri, S., DeFries, R. S., Liang, S., & Yang, K. (2000). Beware of per-pixel characterization of land-cover. *International Journal of Remote Sensing, 21*(4), 839-843. doi:https://doi.org/10.1080/014311600210641

Travers, S. K., Eldridge, D. J., Koen, T. B., & Soliveres, S. (2012). Animal foraging pit soil enhances the performance of a native grass under stressful conditions. *Plant and Soil, 352*(1-2), 341-351. doi:https://doi.org/10.1007/s11104-011-1000-y

Trimble. (2019). eCognition (Version 9.5.1). Munich, germany: Trimble Germany GmbH.

Turney, C. S. M., Flannery, T. F., Roberts, R. G., Reid, C., Fifield, L. K., Higham, T. F. G., . . . Ogle, a. N. (2008). Late-surviving megafauna in Tasmania, Australia, implicate human involvement in their extinction. *PNAS, 105*(34), 12150-12153. doi:https://doi.org/10.1073/pnas.0801360105

Valentine, L. E., Bretz, M., Ruthrof, K. X., Fisher, R., Hardy, G. E. S. J., & Fleming, P. A. (2016). Scratching beneath the surface: bandicoot bioturbation contributes to ecosystem processes. *Austral Ecology, 42*(3). doi:https://doi.org/10.1111/aec.12428

Valentine, L. E., Bretz, M., Ruthrof, K. X., Fisher, R., Hardy, G. E. S. J., & Fleming, P. A. (2017). Scratching beneath the surface: Bandicoot bioturbation contributes to ecosystem processes. *Austral Ecology, 42*(3), 265-276. doi:https://doi.org/10.1111/aec.12428

Valentine, L. E., Ruthrof, K. X., Fisher, R., Hardy, G. E. S. J., Hobbs, R. J., & Fleming, P. A. (2018). Bioturbation by bandicoots facilitates seedling growth by altering soil properties. *Functional Ecology, 32*(9), 2138-2148. doi:https://doi.org/10.1111/1365-2435.13179

Vapnik, V. N. (1995). *The nature of statistical learning theory* (1st ed.). New York, New York: Springer.

VC Technology. (2019). Litchi (Version 4.14.0). London, England: VC Technology Ltd.

Verdon, S. J., Gibb, H., & Leonard, S. W. J. (2016). Net effects of soil disturbance and herbivory on vegetation by a re-established digging mammal assemblage in arid zone Australia. *Journal of Arid Environments, 133*, 29-36. doi:https://doi.org/10.1016/j.jaridenv.2016.05.008

Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human Domination of Earth's Ecosystems. *Science, 277*(5325), 494-499. doi:https://doi.org/10.1126/science.277.5325.494

Waldner, F., Fritz, S., Gregorio, D. A., Plotnikov, D., Bartalev, S., Kussul, N., . . . Defourny, P. (2016). A Unified Cropland Layer at 250 m for Global Agriculture Monitoring. *Data, 1*(1), 2306-5729. doi:https://doi.org/10.3390/data1010003

Wang, L., Gong, P., & Biging, G. S. (2004). Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *American Society for Photogrammetry and Remote Sensing, 3*(7), 351-357. doi:https://doi.org/10.14358/PERS.70.3.351

Wang, L., Zang, J., Zhang, Q., Niu, Z., Hua, G., & Zheng, N. (2018). Action recognition by an attention-aware temporal weighted convolutional neural network. *Sensors, 18*(7). doi:https://doi.org/10.3390/s18071979

Webster, R., & Oliver, M. A. (2001). *Geostatistics for environmental scientists* (1st ed.). Brisbane, Queensland: John Wiley and Sons.

Whitlock, F. L. (1921). Notes of Dirk Hartog Island and Peron Peninsula, Shark Bay, Western Australia. *Emu - Austral Ornithology, 20*(3), 168-186. doi:https://doi.org/10.1071/MU920168

Wickham, H. (2019a). dplyr: a grammar of data manipulation (Version 0.8.3): R package. Retrieved from http://dplyr.tidyverse.org/

Wickham, H. (2019b). ggplot2: create elegant data visualisations using the grammar of graphics (Version 3.2.1): R package. Retrieved from http://ggplot2.tidyverse.org/

Wickham, H. (2019c). stringr: simple, consistent wrappers for common string operations (Version 1.4.0): R package. Retrieved from http://stringr.tidyverse.org/

Wickham, H. (2019d). tidyr: tidy messy data (Version 1.0.0): R package. Retrieved from https://tidyr.tidyverse.org/

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques* (3rd ed.). Burlington, Massachusetts: Morgan Kaufmann.

Woinarski, J. C. Z., Burbidge, A. A., & Harrison, P. L. (2015). Ongoing unraveling of a continental fauna: Decline and extinction of Australian mammals since European settlement. *PNAS, 112*(15), 4531-4540. doi:https://doi.org/10.1073/pnas.1417301112

Wright, M. N. (2019). ranger: a fast implementation of Random Forests (Version 0.11.2): R package. Retrieved from https://github.com/imbs-hl/ranger

Wu, Q., Zhong, R., Zhao, W., Fu, H., & Song, K. (2017). A comparison of pixel-based decision tree and object-based Support Vector Machine methods for land-cover classification based on aerial images and airborne lidar data. *International Journal of Remote Sensing, 38*(23), 7176-7195. doi:https://doi.org/10.1080/01431161.2017.1371864

Xiao, B., Sun, F., Hu, K., & Kidrond, G. J. (2019). Biocrusts reduce surface soil infiltrability and impede soil water infiltration under tension and ponding conditions in dryland ecosystem. *Journal of Hydrology, 568*, 792-802. doi:https://doi.org/10.1016/j.jhydrol.2018.11.051

Yan, G., Mas, J. F., Maathuis, B. H. P., Xiangmin, Z., & Dijk, P. M. V. (2007). Comparison of pixel-based and object-oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China. *International Journal of Remote Sensing, 27*(18), 4039-4055. doi:https://doi.org/10.1080/01431160600702632

Yang, L., Mansaray, L. R., Huang, J., & Wang, L. (2019). Optimal segmentation scale parameter, feature subset and classification algorithm for Geographic Object-Based crop recognition using multisource satellite imagery. *Remote Sensing, 11*(5), 514. doi:http://doi.org/10.3390/rs11050514

Ye, S., Pontius, R. G., & Rakshit, R. (2018). A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS Journal of Photogrammetry and Remote Sensing, 141*, 137-147. doi:https://doi.org/10.1016/j.isprsjprs.2018.04.002

Yong, D. L., Barton, P. S., Ikin, K., Evans, M. J., Crane, M., Okada, S., . . . Lindenmayer, D. B. (2018). Cross-taxonomic surrogates for biodiversity conservation in human modified landscapes – a multi-taxa approach. *Biological Conservation, 224*, 336-346. doi:https://doi.org/10.1016/j.biocon.2018.06.008

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., & Schirokauer, D. (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *American Society for Photogrammetry and Remote Sensing, 72*(7), 799-811. doi:https://doi.org/10.14358/PERS.72.7.799

Zhang, C. (2016). Multiscale quantification of urban composition from EO-1/Hyperion data using object-based spectral unmixing. *International Journal of Applied Earth Observation and Geoinformation, 47*, 153-162. doi:https://doi.org/10.1016/j.jag.2016.01.002

Zhang, C., Denka, S., Cooper, H., & Mishra, D. R. (2018). Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data. *Remote Sensing of Environment, 204*, 366-379. doi:http://doi.org/10.1016/j.rse.2017.10.018

Zhang, L., Jia, K., Li, X., Yuan, Q., & Zhao, X. (2014). Multi-scale segmentation approach for object-based land-cover classification using high-resolution imagery. *Remote Sensing Letters, 5*(1). doi:https://doi.org/10.1080/2150704X.2013.875235

Zhang, X., Xiao, P., & Feng, X. (2012). An unsupervised evaluation method for remotely sensed imagery segmentation. *IEEE, 9*(2), 156-160. doi:https://doi.org/10.1109/LGRS.2011.2163056

Zhang, X., Xiao, P., Song, X., & She, J. (2013). Boundary-constrained multi-scale segmentation method for remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing, 78*, 15-25. doi:https://doi.org/10.1016/j.isprsjprs.2013.01.002

Zhang, Y., & Maxwell, T. (2006). *A fuzzy logic approach to supervised segmentation for object oriented classification*. Paper presented at the ASPRS, Reno, Nevada.

Zhou, Z., Ma, L., Fu, T., Zhang, G., Yao, M., & Manchun. (2018). Change detection in coral reef environment using high-resolution images: comparison of object-based and pixel-based paradigms. *ISPRS International Journal of Geo-Information, 7*(11), 441. doi:https://doi.org/10.3390/ijgi7110441

Zhu, H., Cai, L., Liu, H., & Huang, W. (2016). Information extraction of high resolution remote sensing images based on the calculation of optimal segmentation parameters. *PLoS One, 11*(6), e0158585. doi:https://doi.org/10.1371/journal.pone.0158585

8. Appendix

This section details the additional processing information and results that is supplementary to the data outlined in previous chapters.

8.1. R code

```
library(dplyr)
library(stringr)
library(ggplot2)
library(sf)
library(tidyr)
library(raster)
library(PerformanceAnalytics)
library(xlsx)
library(sciplot)

gpsdir <- "C:\\Users\\Lucy\\Desktop\\Curtin University\\Masters\\Dirk
Hartog\\DirkHartogData\\plant_gps\\" # data location

dhi_plantboundaries <- st_read(paste0(gpsdir,
"PlantBoundarie_PlantIdAppend_20190912.shp"), quiet = TRUE) # input digitised
plant boundary data across the 3 field days

dhi_gps <- dhi_plantboundaries %>%
  dplyr::select(Name) %>% # select only the "name", removing others
  filter(str_detect(Name, "^x")) # filter for exclusion plot identifiers

colnames(dhi_gps)[1] <- "Waypoint"  # rename the name field so datasets can
be easily joined

ex01csv <- read.csv(paste0(gpsdir, "ex01.csv")) # load in field data for
exclusion plot 1
ex06csv <- read.csv(paste0(gpsdir, "ex06.csv")) # load in field data for
exclusion plot 6

ex01 <- dhi_gps %>% left_join(ex01csv, by = "Waypoint") %>% # join field data
with plant boundaries
  filter(str_detect(Waypoint, "^x01"))  # filter for ex01 records

ex01 <- na.omit(ex01) # remove NA records

ex06 <- left_join(dhi_gps, ex06csv,  by = "Waypoint")%>% # join field data
with plant  boundaries
  filter(str_detect(Waypoint, "^x06")) # filter for ex06 records

# exclusion plot 1
# group by and summarise functions - statistics per species can be calculated
ex01g <- ex01 %>% group_by(SpeciesNotes) %>%
  summarise(n = n())

# plot the number of records by species
ggplot(ex01g, aes(SpeciesNotes, n)) +
  geom_bar(stat = "identity") +
```

```
  theme(axis.text.x = element_text(angle = 90, hjust = 1))+
  labs(y = "Number of individuals", x = "Species")

data(ex01g)
View(ex01g) # view data for exact species counts

# select species where samples total >= 5
ex01 <- subset(ex01, SpeciesNotes!="Westringia dampieri")
ex01 <- subset(ex01, SpeciesNotes!="Santalum acuminatum")

# exclusion plot 6
# group by and summarise functions - statistics per species can be calculated
ex06g <- ex06 %>% group_by(SpeciesNotes) %>%
  summarise(n = n())

# plot the number of records by species
ggplot(ex06g, aes(SpeciesNotes, n)) +
  geom_bar(stat = "identity") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  labs(y = "Number of individuals", x = "Species")

data(ex06g)
View(ex06g) # view data for exact species counts

# select species where samples total >= 5
ex01 <- subset(ex01, SpeciesNotes!="Westringia dampieri")
ex01 <- subset(ex01, SpeciesNotes!="Santalum acuminatum")

# segment data location
ecogdir <- gpsdir <- "C:\\Users\\Lucy\\Desktop\\Curtin University\\Dirk
Hartog\\DirkHartogData\\segments\\results_20191016\\"

# BOX-PLOTTING VARIABLES
# exclusion plot 1
ecog01 <- st_read(paste0(ecogdir, "Site_01b.v50.shp"), quiet =  TRUE)  # read
in segmentation results

ex01 <- st_transform(ex01, crs(ecog01)) # reproject field data to match
segments (GDA 1994 MGA 49)

ex01st <- st_intersection(ex01, ecog01) # intersect the field data and
segments
st_geometry(ex01st) <- NULL # remove geometry

ex01st <- dplyr::select(ex01st, SpeciesNotes, Mean_blue, Mean_green,
Mean_red, Med_gla, Max_gla, Mean_gla, Med_nsm, Min_nsm, Mean_nsm, Max_nsm,
quan90nsm, quan90gla, Roundness, GLCM_Contr, GLCM_Corre, GLCM_Entro,
GLCM_Homog, GLCM_Mean_, LengthWidt, Area, Compactnes) # select variables for
analysis

for(i in 2:ncol(ex01st)){
ex01st[,i] <- as.numeric(ex01st[,i])
} # for loop to set data up as numerical for statistical analysis

# segment totals
ex01seg <- ex01st %>% group_by(SpeciesNotes) %>% # using the group_by and
summarise functions stats per group can be calculated
```

```
  summarise(n = n())

# view total segments per species
data(ex01seg)
View(ex01seg)

ex01box1 <- gather(ex01st, "measure", "value", 2:22) # change the data to
"long" format, this allows you to use the facet wrap in ggplot

# plot
ggplot(ex01box1, aes(x = SpeciesNotes, y = value)) +
        geom_boxplot() +
        facet_wrap(~measure , scales="free")  +
        theme(axis.text.x = element_text(angle = 90, hjust = 1))

# exclusion plot 6
ecog06 <- st_read(paste0(ecogdir, "Site_07.v50.shp"), quiet =  TRUE) # read
in segmentation results

ex06 <- st_transform(ex06, crs(ecog06)) # reproject field data to match
segments (GDA 1994 MGA 49)

ex06st <- st_intersection(ex06, ecog06) # intersect the field data and
segments
st_geometry(ex06st) <- NULL # remove geometry

ex06st <- dplyr::select(ex06st, SpeciesNotes, Mean_blue, Mean_green,
Mean_red, Med_gla, Max_gla, Mean_gla, Med_nsm, Min_nsm, Mean_nsm, Max_nsm,
quan90nsm, quan90gla, Roundness, GLCM_Contr, GLCM_Corre, GLCM_Entro,
GLCM_Homog, GLCM_Mean_, LengthWidt, Area, Compactnes) # select variables for
analysis

for(i in 2:ncol(ex06st)){
ex06st[,i] <- as.numeric(ex06st[,i])
} # for loop to set data up as numerical for statistical analysis

# segment totals
ex06seg <- ex06st %>% group_by(SpeciesNotes) %>% # using the group_by and
summarise functions stats per group can be calculated
  summarise(n = n())

# view total segments per species
data(ex06seg)
View(ex06seg)

ex06box1 <- gather(ex06st, "measure", "value", 2:22) # change the data to
"long" format, this allows you to use the facet wrap in ggplot

# plot
ggplot(ex06box1, aes(x = SpeciesNotes, y = value)) +
        geom_boxplot() +
        facet_wrap(~measure , scales="free")  +
        theme(axis.text.x = element_text(angle = 90, hjust = 1))

# STATISTICAL TESTING

# Exclusion Plot 1
```

```
# set up response variable as factor
ex01st$SpeciesNotes <- as.factor(ex01st$SpeciesNotes)
str(ex01st) # check data to ensure factor has been applied correctly

# standard deviations
sd(ex01st$Mean_blue)
sd(ex01st$Mean_green)
sd(ex01st$Mean_red)
sd(ex01st$Med_gla)
sd(ex01st$Max_gla)
sd(ex01st$Mean_gla)
sd(ex01st$Med_nsm)
sd(ex01st$Min_nsm)
sd(ex01st$Mean_nsm)
sd(ex01st$Max_nsm)
sd(ex01st$quan90nsm)
sd(ex01st$quan90gla)
sd(ex01st$Roundness)
sd(ex01st$GLCM_Contr)
sd(ex01st$GLCM_Corre)
sd(ex01st$GLCM_Entro)
sd(ex01st$GLCM_Homog)
sd(ex01st$GLCM_Mean_)
sd(ex01st$LengthWidt)
sd(ex01st$Area)
sd(ex01st$Compactnes)

# standard error
se(ex01st$Mean_blue)
se(ex01st$Mean_green)
se(ex01st$Mean_red)
se(ex01st$Med_gla)
se(ex01st$Max_gla)
se(ex01st$Mean_gla)
se(ex01st$Med_nsm)
se(ex01st$Min_nsm)
se(ex01st$Mean_nsm)
se(ex01st$Max_nsm)
se(ex01st$quan90nsm)
se(ex01st$quan90gla)
se(ex01st$Roundness)
se(ex01st$GLCM_Contr)
se(ex01st$GLCM_Corre)
se(ex01st$GLCM_Entro)
se(ex01st$GLCM_Homog)
se(ex01st$GLCM_Mean_)
se(ex01st$LengthWidt)
se(ex01st$Area)
se(ex01st$Compactnes)

# skewness value calculation
skewness(ex01st$Mean_red)
skewness(ex01st$Mean_gla)
skewness(ex01st$Med_nsm)
skewness(ex01st$GLCM_Homog)
skewness(ex01st$GLCM_Entro)
```

```
skewness(ex01st$Compactnes)

attach(ex01st) # select data source
y1 <- cbind(Mean_blue, Mean_green, Mean_red, Med_gla, Max_gla, Mean_gla,
Med_nsm, Min_nsm, Mean_nsm, Max_nsm, quan90nsm, quan90gla, Roundness,
GLCM_Contr, GLCM_Corre, GLCM_Entro, GLCM_Homog, GLCM_Mean_, LengthWidt, Area,
Compactnes) # bind explanatory variables

summary(y1) # summary statistics

# skewness value calculation
skewness(ex01st$Mean_red)
skewness(ex01st$Mean_gla)
skewness(ex01st$Med_nsm)
skewness(ex01st$Min_nsm)
skewness(ex01st$GLCM_Homog)
skewness(ex01st$GLCM_Entro)
skewness(ex01st$Compactnes)
skewness(ex01st$Area)

chart.Correlation(y1) # test MANOVA assumtions
# no correlations > 0.6
# require roughly normal distributions
# require roughly linear relationships

# data transformations
ex01st$Mean_gla_pos <- -ex01st$Mean_gla # reflect negatively skewed variables
ex01st$Mean_gla_pos <- ex01st$Mean_gla_pos + (max(ex01st$Mean_gla) + 1) #
create positive Mean_gla values for transformation

ex01st$GLCM_Entro_pos <- -ex01st$GLCM_Entro # reflect negatively skewed
variables
ex01st$GLCM_Entro_pos <- ex01st$GLCM_Entro_pos + (max(ex01st$GLCM_Entro) + 1)
#create positive GLCM_Entro values for transformation

# histograms for determining if positive skew amended and presence of zero
values
hist(ex01st$Mean_red,  main = "a)", col="gray")
hist(ex01st$Mean_gla_pos,  main = "b)", col="gray")
hist(ex01st$Med_nsm,  main = "c)", col="gray")
hist(ex01st$Min_nsm,  main = "d)", col="gray")
hist(ex01st$GLCM_Homog,  main = "e)", col="gray")
hist(ex01st$GLCM_Entro_pos,  main = "f)", col="gray")
hist(ex01st$Area,  main = "g)", col="gray")

ex01st$Mean_red_TranF <- ex01st$Mean_red # set up sqrt transformed Mean_red
variable
ex01st$Mean_gla_TranF <- ex01st$Mean_gla_pos # set up sqrt transformed
Mean_gla variable
ex01st$Med_nsm_TranF <- ex01st$Med_nsm # set up log transformed Med_nsm
variable
ex01st$Min_nsm_TranF <- ex01st$Min_nsm # set up log transformed Min_nsm
variable
ex01st$GLCM_Homog_TranF <- ex01st$GLCM_Homog # set up log transformed
GLCM_Homog variable
ex01st$GLCM_Entro_TranF <- ex01st$GLCM_Entro_pos # set up log transformed
GLCM_Entro variable
```

```
ex01st$Area_TranF <- ex01st$Area # set up log transformed Area variable

ex01st$Med_nsm_TranF <- ex01st$Med_nsm_TranF + 1 # apply constant to
Med_nsm_Tranf values for future log transform
ex01st$Min_nsm_TranF <- ex01st$Min_nsm_TranF + 1 # apply constant to
Min_nsm_TranF values for future log transform
ex01st$GLCM_Homog_TranF <- ex01st$GLCM_Homog_TranF + 1 # apply constant to
GLCM_Homog_TranF values for future log transform
ex01st$Area_TranF <- ex01st$Area_TranF + 1 # apply constant to Area_TranF
values for future log transform

ex01st$Mean_red_TranF <- sqrt(ex01st$Mean_red_TranF) # square-root
transformation
ex01st$Mean_gla_TranF <- sqrt(ex01st$Mean_gla_TranF) # square-root
transformation
ex01st$Med_nsm_TranF <- log(ex01st$Med_nsm_TranF) # logarithmic
transformation
ex01st$Min_nsm_TranF <- log(ex01st$Min_nsm_TranF) # logarithmic
transformation
ex01st$GLCM_Homog_TranF <- log(ex01st$GLCM_Homog_TranF) #logarithmic
transformation
ex01st$GLCM_Entro_TranF <- log(ex01st$GLCM_Entro_TranF) #logarithmic
transformation
ex01st$Area_tranF <- log(ex01st$Area_TranF) #logarithmic transformation

ex01st$Med_nsm_TranF[is.infinite(ex01st$Med_nsm_TranF)] <- 0 # ensure log has
worked on 1 values
ex01st$Min_nsm_TranF[is.infinite(ex01st$Min_nsm_TranF)] <- 0 # ensure log has
worked on 1 values

# re-applying previous assumption tests
skewness(ex01st$Mean_red_TranF)
skewness(ex01st$Mean_gla_TranF)
skewness(ex01st$Med_nsm_TranF)
skewness(ex01st$Min_nsm_TranF)
skewness(ex01st$GLCM_Homog_TranF)
skewness(ex01st$GLCM_Entro_TranF)
skewness(ex01st$Area_TranF)

# histograms of transformed data
hist(ex01st$Mean_red_TranF,  main = "a)", col="gray")
hist(ex01st$Mean_gla_TranF,  main = "b)", col="gray")
hist(ex01st$Med_nsm_TranF,  main = "c)", col="gray")
hist(ex01st$Min_nsm_TranF,  main = "d)", col="gray")
hist(ex01st$GLCM_Homog_TranF,  main = "e)", col="gray")
hist(ex01st$GLCM_Entro_TranF,  main = "f)", col="gray")
hist(ex01st$Area_TranF,  main = "g)", col="gray")

hist(ex06st$Med_nsm_TranF,  main = "a)", col="gray")
hist(ex06st$Min_nsm_TranF,  main = "b)", col="gray")
hist(ex06st$GLCM_Homog_TranF,  main = "c)", col="gray")
hist(ex06st$GLCM_Contr_TranF,  main = "d)", col="gray")
hist(ex06st$Area_TranF,  main = "e)", col="gray")

# further assumption testing on the refined dataset
attach(ex01st)
```

```
FinalTestx01 <- cbind(Mean_red_TranF, Mean_gla_TranF, Med_nsm_TranF,
Min_nsm_TranF, GLCM_Entro_TranF, Compactnes) # bind data set

chart.Correlation(FinalTestx01) # final correlation matrix
mvoutlier::aq.plot(FinalTestx01, quan = 1) # check for outliers

write.xlsx(ex01st, file="C:\\Users\\Lucy\\Desktop\\Curtin
#University\\Masters\\Dirk Hartog\\DirkHartogData\\rawMANOVAdata.xlsx",
sheetName = "x01", append = FALSE) # raw and transformed data table export to
excel

# Exclusion Plot 6

# set up response variable as factor
ex06st$SpeciesNotes <- as.factor(ex06st$SpeciesNotes)

# standard deviations
sd(ex06st$Mean_blue)
sd(ex06st$Mean_green)
sd(ex06st$Mean_red)
sd(ex06st$Med_gla)
sd(ex06st$Max_gla)
sd(ex06st$Mean_gla)
sd(ex06st$Med_nsm)
sd(ex06st$Min_nsm)
sd(ex06st$Mean_nsm)
sd(ex06st$Max_nsm)
sd(ex06st$quan90nsm)
sd(ex06st$quan90gla)
sd(ex06st$Roundness)
sd(ex06st$GLCM_Contr)
sd(ex06st$GLCM_Corre)
sd(ex06st$GLCM_Entro)
sd(ex06st$GLCM_Homog)
sd(ex06st$GLCM_Mean_)
sd(ex06st$LengthWidt)
sd(ex06st$Area)
sd(ex06st$Compactnes)

# standard error
se(ex06st$Mean_blue)
se(ex06st$Mean_green)
se(ex06st$Mean_red)
se(ex06st$Med_gla)
se(ex06st$Max_gla)
se(ex06st$Mean_gla)
se(ex06st$Med_nsm)
se(ex06st$Min_nsm)
se(ex06st$Mean_nsm)
se(ex06st$Max_nsm)
se(ex06st$quan90nsm)
se(ex06st$quan90gla)
se(ex06st$Roundness)
se(ex06st$GLCM_Contr)
se(ex06st$GLCM_Corre)
se(ex06st$GLCM_Entro)
se(ex06st$GLCM_Homog)
```

```
se(ex06st$GLCM_Mean_)
se(ex06st$LengthWidt)
se(ex06st$Area)
se(ex06st$Compactnes)

attach(ex06st) # select data source
y6 <- cbind(Mean_blue, Mean_green, Mean_red, Med_gla, Max_gla, Mean_gla,
Med_nsm, Min_nsm, Mean_nsm, Max_nsm, quan90nsm, quan90gla, Roundness,
GLCM_Contr, GLCM_Corre, GLCM_Entro, GLCM_Homog, GLCM_Mean_, LengthWidt, Area,
Compactnes) # bind explanatory variables

summary(y6) # summary statistics

chart.Correlation(y6) # test MANOVA assumtions
# no correlations > 0.6
# require roughly normal distributions
# require roughly linear relationships

# skewness value calculation
skewness(ex06st$Mean_red)
skewness(ex06st$Max_gla)
skewness(ex06st$Med_nsm)
skewness(ex06st$Min_nsm)
skewness(ex06st$GLCM_Homog)
skewness(ex06st$GLCM_Contr)
skewness(ex06st$Compactnes)
skewness(ex06st$Area)

# data transformations
# histograms for determining presence of zero values
hist(ex06st$Med_nsm,  main = "a)", col="gray")
hist(ex06st$Min_nsm,  main = "b)", col="gray")
hist(ex06st$GLCM_Homog,  main = "c)", col="gray")
hist(ex06st$GLCM_Contr,  main = "d)", col="gray")
hist(ex06st$Area,  main = "e)", col="gray")

ex06st$Med_nsm_TranF <- ex06st$Med_nsm # set up square root transformed
Med_nsm variable
ex06st$Min_nsm_TranF <- ex06st$Min_nsm # set up square root transformed
Min_nsm variable
ex06st$GLCM_Homog_TranF <- ex06st$GLCM_Homog # set up log transformed
GLCM_Homog variable
ex06st$GLCM_Contr_TranF <- ex06st$GLCM_Contr # set up log transformed
GLCM_Entro variable
ex06st$Area_TranF <- ex06st$Area # set up log transformed Area variable

ex06st$Med_nsm_TranF <- sqrt(ex06st$Med_nsm_TranF) # square-root
transformation
ex06st$Min_nsm_TranF <- sqrt(ex06st$Min_nsm_TranF) # square-root
transformation
ex06st$GLCM_Homog_TranF <- log(ex06st$GLCM_Homog_TranF) #logarithmic
transformation
ex06st$GLCM_Contr_TranF <- log(ex06st$GLCM_Contr_TranF) #logarithmic
transformation
ex06st$Area_TranF <- log(ex06st$Area_TranF) #logarithmic transformation

# re-applying previous assumption tests
```

```
skewness(ex06st$Med_nsm_TranF)
skewness(ex06st$Min_nsm_TranF)
skewness(ex06st$GLCM_Homog_TranF)
skewness(ex06st$GLCM_Contr_TranF)
skewness(ex06st$Area_TranF)

# further assumption testing on the refined dataset
attach(ex06st)
FinalTestx06 <- cbind(Mean_red, Max_gla, Med_nsm_TranF, Min_nsm_TranF,
GLCM_Contr_TranF, Compactnes, Area_TranF) # bind data set

chart.Correlation(FinalTestx06) # final correlation matrix
mvoutlier::aq.plot(FinalTestx06, quan = 1) # check for outliers

ex06boxFinal <- gather(ex06st, "measure", "value", 26:27) # change the data
to "long" format, this allows you to use the facet wrap in ggplot

# recreate box plot for comparisons of Area_TranF and GLCM_Contr_TranF
ggplot(ex06boxFinal, aes(x = SpeciesNotes, y = value)) +
        geom_boxplot() +
        facet_wrap(~measure , scales="free")  +
        theme(axis.text.x = element_text(angle = 90, hjust = 1))


write.xlsx(ex06st, file="C:\\Users\\Lucy\\Desktop\\Curtin
University\\Masters\\Dirk Hartog\\DirkHartogData\\rawMANOVAdata.xlsx",
sheetName = "x06", append = TRUE) # raw and transformed data table export to
excel

# MANOVA TESTS
attach(ex01st)
MANOVA_Data01 <- cbind(Mean_red_TranF, Mean_gla_TranF, Med_nsm_TranF,
GLCM_Entro_TranF, Compactnes) # bind data set for MANOVA
RF.MANOVA1 <- manova(MANOVA_Data01 ~ ex01st$SpeciesNotes) # set up MANOVA
interaction terms
summary(RF.MANOVA1) # MANOVA test for significance
summary.aov(RF.MANOVA1) # post-hoc univariate test

attach(ex06st)
MANOVA_Data06 <- cbind(Mean_red, Max_gla, Med_nsm_TranF, GLCM_Contr_TranF,
Compactnes) # bind data set for MANOVA
RF.MANOVA6 <- manova(MANOVA_Data06 ~ ex06st$SpeciesNotes) # set up MANOVA
interaction terms
summary(RF.MANOVA6) # MANOVA test for significance
summary.aov(RF.MANOVA6) # post-hoc univariate test

# Random Forest (RF) run over 20, 50, and 100 segmentation scales for x01 and
x06 across all original variables. Scale 50 was tested for all variables,
selected raw variables and selected variable set with transformed data.

ex01st <- droplevels(ex01st) # remove factorisation of values
ex06st <- droplevels(ex06st) # remove factorisation of values

ex01All <- dplyr::select(ex01st, -Mean_gla_pos, -GLCM_Entro_pos, -
Mean_red_TranF, - Mean_gla_TranF, -Med_nsm_TranF, -Min_nsm_TranF, -
GLCM_Homog_TranF, -GLCM_Entro_TranF, -Area_TranF) # original ex01 variable
capture
```

```
ex01TFSelect <- dplyr::select(ex01st,  SpeciesNotes, Mean_red_TranF,
Mean_gla_TranF, Med_nsm_TranF, GLCM_Entro_TranF, Compactnes) # example
transformed variable set for x01

ex01RawSelect <- dplyr::select(ex01st, SpeciesNotes, Mean_red, Mean_gla,
Med_nsm, GLCM_Entro, Compactnes) # example untransformed variable set for x01

ex06All <- dplyr::select(ex06st, -Med_nsm_TranF, -Min_nsm_TranF, -
GLCM_Contr_TranF, -GLCM_Homog_TranF, -Area_TranF)  # original ex06 variable
capture

ex06RawSelect <- dplyr::select(ex06st,  SpeciesNotes, Mean_red, Max_gla,
Med_nsm, GLCM_Contr, Compactnes) # example untransformed variable set for x06

ex06TFSelect <- dplyr::select(ex06st,  SpeciesNotes, Mean_red, Max_gla,
Med_nsm_TranF, GLCM_Contr_TranF, Compactnes) # example transformed variable
set for x06

# RF model example - not all scenarios shown as model was adapted per
iteration
# RF iteration 1
mtry <- seq(from = 2, to = ncol(ex01All)-1, by = 2) # set up mtry value,
iterating through variables by 1

tunegrid <- expand.grid(.mtry=mtry, # set up tune grid per mtry value, with
split rule extra trees
                        .splitrule = c( "extratrees"),#,
                        .min.node.size = 1) # may have only one item in the
final vote
# increasing min node size threshold will apply additional tree splits to
ensure minimum class size is met

ranger_model <- train(
  SpeciesNotes ~., # variable used to train model
  tuneGrid = tunegrid,
  data = ex01All, method = "ranger", # random forest model from the ranger
package
  trControl = trainControl(method = "cv",
                           number = 10, # number of folds in the trainings
data
                           verboseIter = TRUE)) # do not print progress
updates

max(ranger_model$results["Accuracy"], na.rm = TRUE)
max(ranger_model$results["Kappa"], na.rm = TRUE)

# RF iteration 2
mtry <- seq(from = 2, to = ncol(ex01All)-1, by = 2) # set up mtry value,
iterating through variables by 1

tunegrid <- expand.grid(.mtry=mtry, # set up tune grid per mtry value, with
split rule extra trees
                        .splitrule = c( "extratrees"),#,
                        .min.node.size = 1) # may have only one item in the
final vote
```

```
# increasing min node size threshold will apply additional tree splits to
ensure minimum class size is met

ranger_model <- train(
  SpeciesNotes ~., # variable used to train model
  tuneGrid = tunegrid,
  data = ex01All, method = "ranger", # random forest model from the ranger
package
  trControl = trainControl(method = "cv",
                           number = 10, # number of folds in the trainings
data
                           verboseIter = TRUE)) # do not print progress
updates

max(ranger_model$results["Accuracy"], na.rm = TRUE)
max(ranger_model$results["Kappa"], na.rm = TRUE)

# RF iteration 3
mtry <- seq(from = 2, to = ncol(ex01All)-1, by = 2) # set up mtry value,
iterating through variables by 1

tunegrid <- expand.grid(.mtry=mtry, # set up tune grid per mtry value, with
split rule extra trees
                        .splitrule = c( "extratrees"),#,
                        .min.node.size = 1) # may have only one item in the
final vote
# increasing min node size threshold will apply additional tree splits to
ensure minimum class size is met

ranger_model <- train(
  SpeciesNotes ~., # variable used to train model
  tuneGrid = tunegrid,
  data = ex01All, method = "ranger", # random forest model from the ranger
package
  trControl = trainControl(method = "cv",
                           number = 10, # number of folds in the trainings
data
                           verboseIter = TRUE)) # do not print progress
updates

max(ranger_model$results["Accuracy"], na.rm = TRUE)
max(ranger_model$results["Kappa"], na.rm = TRUE)

# RF iteration 4
mtry <- seq(from = 2, to = ncol(ex01All)-1, by = 2) # set up mtry value,
iterating through variables by 1

tunegrid <- expand.grid(.mtry=mtry, # set up tune grid per mtry value, with
split rule extra trees
                        .splitrule = c( "extratrees"),#,
                        .min.node.size = 1) # may have only one item in the
final vote
# increasing min node size threshold will apply additional tree splits to
ensure minimum class size is met

ranger_model <- train(
  SpeciesNotes ~., # variable used to train model
```

```
  tuneGrid = tunegrid,
  data = ex01All, method = "ranger", # random forest model from the ranger
package
  trControl = trainControl(method = "cv",
                           number = 10, # number of folds in the trainings
data
                           verboseIter = TRUE)) # do not print progress
updates

max(ranger_model$results["Accuracy"], na.rm = TRUE)
max(ranger_model$results["Kappa"], na.rm = TRUE)

# RF iteration 5
mtry <- seq(from = 2, to = ncol(ex01All)-1, by = 2) # set up mtry value,
iterating through variables by 1

tunegrid <- expand.grid(.mtry=mtry, # set up tune grid per mtry value, with
split rule extra trees
                        .splitrule = c( "extratrees"),#,
                        .min.node.size = 1) # may have only one item in the
final vote
# increasing min node size threshold will apply additional tree splits to
ensure minimum class size is met

ranger_model <- train(
  SpeciesNotes ~., # variable used to train model
  tuneGrid = tunegrid,
  data = ex01All, method = "ranger", # random forest model from the ranger
package
  trControl = trainControl(method = "cv",
                           number = 10, # number of folds in the trainings
data
                           verboseIter = TRUE)) # do not print progress
updates

max(ranger_model$results["Accuracy"], na.rm = TRUE)
max(ranger_model$results["Kappa"], na.rm = TRUE)

# Continued for 5 more iterations. Total of 10 iterations per scenario.

saveRDS(ranger_model, paste0(ecogdir, "ex01_ranger_", Sys.Date(), ".rds")) #
set-up location and information

ranger_model <- readRDS(paste0(ecogdir, "ex01_ranger_2019-11-10.rds"))

ecog01 <- st_read(paste0(ecogdir, "\\Site_01b.v50.shp"), quiet =  TRUE) #
read in results from ecog

tmp <- ecog01 # temporary variable for saving out classified segments

st_geometry(tmp) <- NULL # remove geometry

i <- 1 # for loop to determine variable set
for(i in 1:ncol(tmp)){
tmp[,i] <- as.numeric(tmp[,i])
}
```

```
ecog01$predict <- predict(ranger_model, tmp, type = "raw") # create class
prediction column

st_write(ecog01, paste0(ecogdir, "\\Site_01b.v50_predict.shp"), quiet = TRUE)
# export classified shapefile
```

8.2. Sampling and segmentation outcomes

Sampling effort for both exclusion plot sites quantified per total species may be found in Table 8.1. Descriptive statistics including range, median, mean, standard deviation, and standard error were derived for each variable per exclusion plot (Table 8.2-3).

Table 8.1. Sample counts per plant species/ground for x01 and x06.

Class	Total number of samples	
	x01	x06
Acacia ligulata	25	21
Acanthocarpus preissii	17	-
Alyogyne cuneiformis	25	-
Atriplex vesicaria	27	20
Cenchrus ciliaris	22	-
Exocarpus aphyllus	19	-
Ground	20	20
Santalum acuminatum	1	-
Scaevola spinescens	-	21
Threlkeldia diffusa	13	-
Triodia plurinervata	16	20
Westringia dampieri	4	-
X01-008	10	-

Table 8.2. Descriptive statistics for the untransformed segment variables from x01.

Variable	Range	Median	Mean	Standard Deviation (±)	Standard Error (±)
Mean_blue	4.38-213.97	85.59	97.19	52.53	0.45
Mean_green	8.69-223.25	105.26	113.52	51.64	0.44
Mean_red	4.33-241.68	99.78	112.5	59.28	0.51
Med_gla	21-7,657	3,878	4,344	2,033.33	17.4
Max_gla	92-12,073	7,832	6,901	3,467.95	29.67
Mean_gla	80-164,934	136,373	117,955	45,197.05	386.74
Med_nsm	0.0-255	0.0	58.25	77.35	0.66
Min_nsm	0.0-255	0.0	39.83	64.61	0.55
Mean_nsm	0.0-255	25.55	60.47	72.65	0.62
Max_nsm	0.0-255	62	82.76	82.97	0.71
quan90nsm	0.0-255	46	74.47	80.52	0.69
quan90gla	27-20,246	13,968	12,577	5,537.03	47.69
Roundness	0.04-4.6	1.28	1.31	0.45	0.0
GLCM_Contr	27-4,520	963.8	1,038.2	433.39	3.71
GLCM_Corre	0.32-0.98	0.8	0.79	0.06	0.0
GLCM_Entro	4.74-9.39	8.24	8.15	0.56	0.0
GLCM_Homog	0.01-0.35	0.04	0.05	0.02	0.0
GLCM_Mean_	36-164,788	86,201	84,530	44,942.05	380.75
LengthWidt	1.0-7.2	1.66	1.79	0.63	0.01
Area	23-13,203	768	1,014	884.3	7.57
Compactness	9-152,635	69,999	71,741	43,180.33	369.48

Table 8.3. Descriptive statistics for the untransformed segment variables from x06.

Variable	Range	Median	Mean	Standard Deviation (±)	Standard Error (±)
Mean_blue	9.72-238.5	108.81	115.45	52.96	0.69
Mean_green	11.03-247.88	125.51	127.44	52.78	0.68
Mean_red	6.35-253.6	124.15	127.63	56.31	0.73
Med_gla	2-7,180	2,704	3,598	1,770.02	22.93
Max_gla	80-10,560	6,247	5,948	2,617.38	33.91
Mean_gla	9-104,782	72,864	67,582	27,947.53	362.1
Med_nsm	0.0-1.55	0.36	0.41	0.38	0.0
Min_nsm	0.0-1.47	0.17	0.3	0.34	0.0
Mean_nsm	0.0-1.53	0.37	0.42	0.37	0.0
Max_nsm	0.0-1.58	0.53	0.53	0.39	0.01
quan90nsm	0.0-1.58	0.48	0.49	0.39	0.01
quan90gla	28-16,763	10,035	9,571	4,289.53	55.58
Roundness	0.18-3.36	1.37	1.39	0.46	0.01
GLCM_Contr	283.7-4,511.5	1,344.5	1,448.8	509.58	6.6
GLCM_Corre	0.49-0.94	0.81	0.8	0.06	0.0
GLCM_Entro	4.48-9.52	8.31	8.24	0.54	0.01
GLCM_Homog	0.01-0.21	0.03	0.04	0.02	0.0
GLCM_Mean_	26-104,772	52,387	52,045	29,906.89	387.49
LengthWidt	1.0-14.62	1.72	1.86	0.71	0.01
Area	40-8,680	702	864.9	650.27	8.42
Compactness	7-98,719	48,496	48,933	29,021.06	376.01

8.3. Data transformations

Due to the volume of segments for x01 (165,935) and x06 (104,806) there was insufficient space to show the complete transformation dataset tables here. The excel file containing raw and transformed segment values is available from the author.

Histograms were derived for the reflected and non-reflected variables before transformation across both exclusion plot sites (Figure 8.1-2). It was determined that the positive skew found across Mean_gla and GLCM_Entro x01 was amended so that further transformations could be applied.

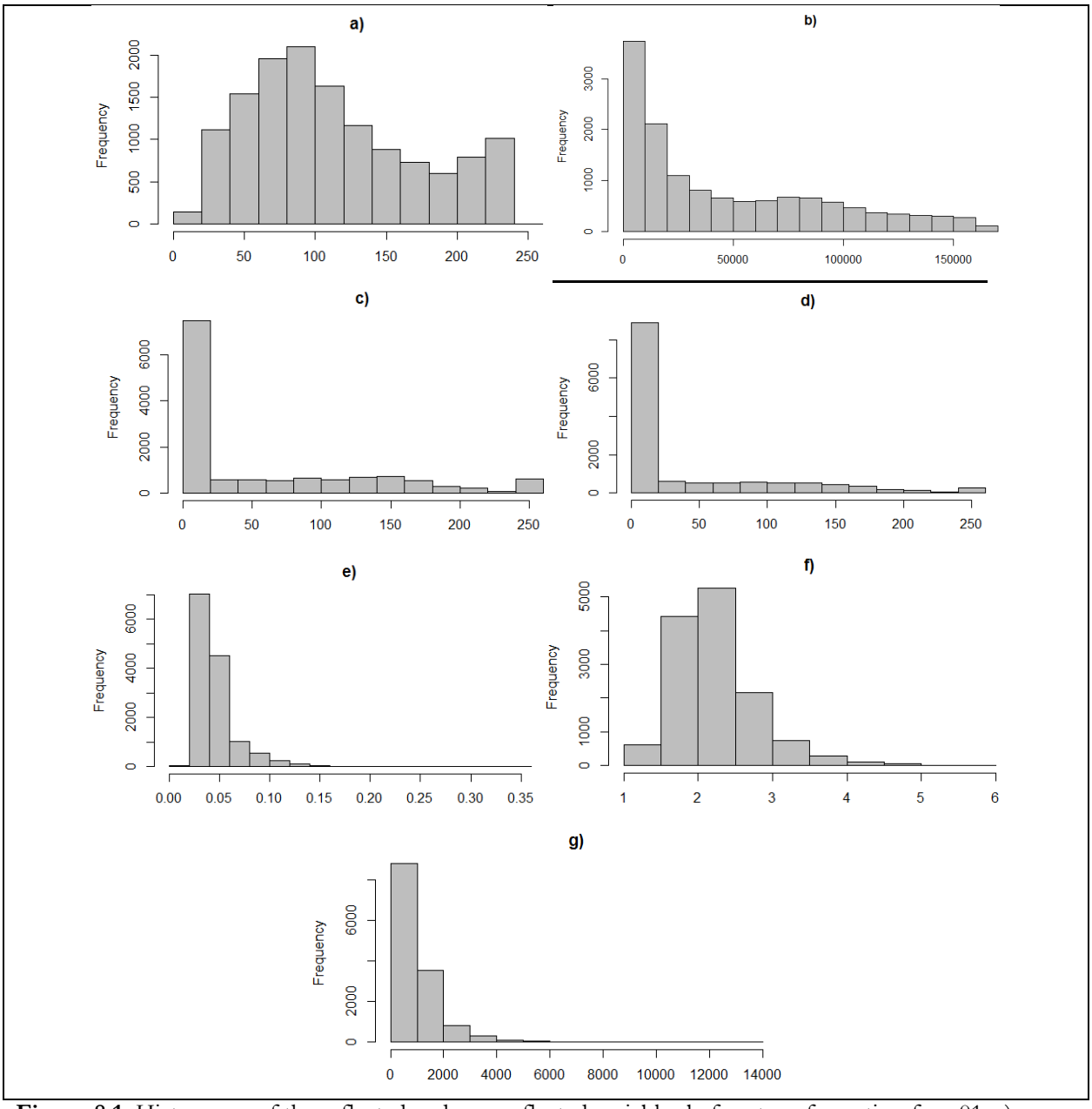

Figure 8.1. Histograms of the reflected and non-reflected variables before transformation for x01. a) Mean_red, b) reflected Mean_gla, c) Med_nsm, d) Min_nsm, e) GLCM_Homog, f) reflected GLCM_Entro, and g) Area.

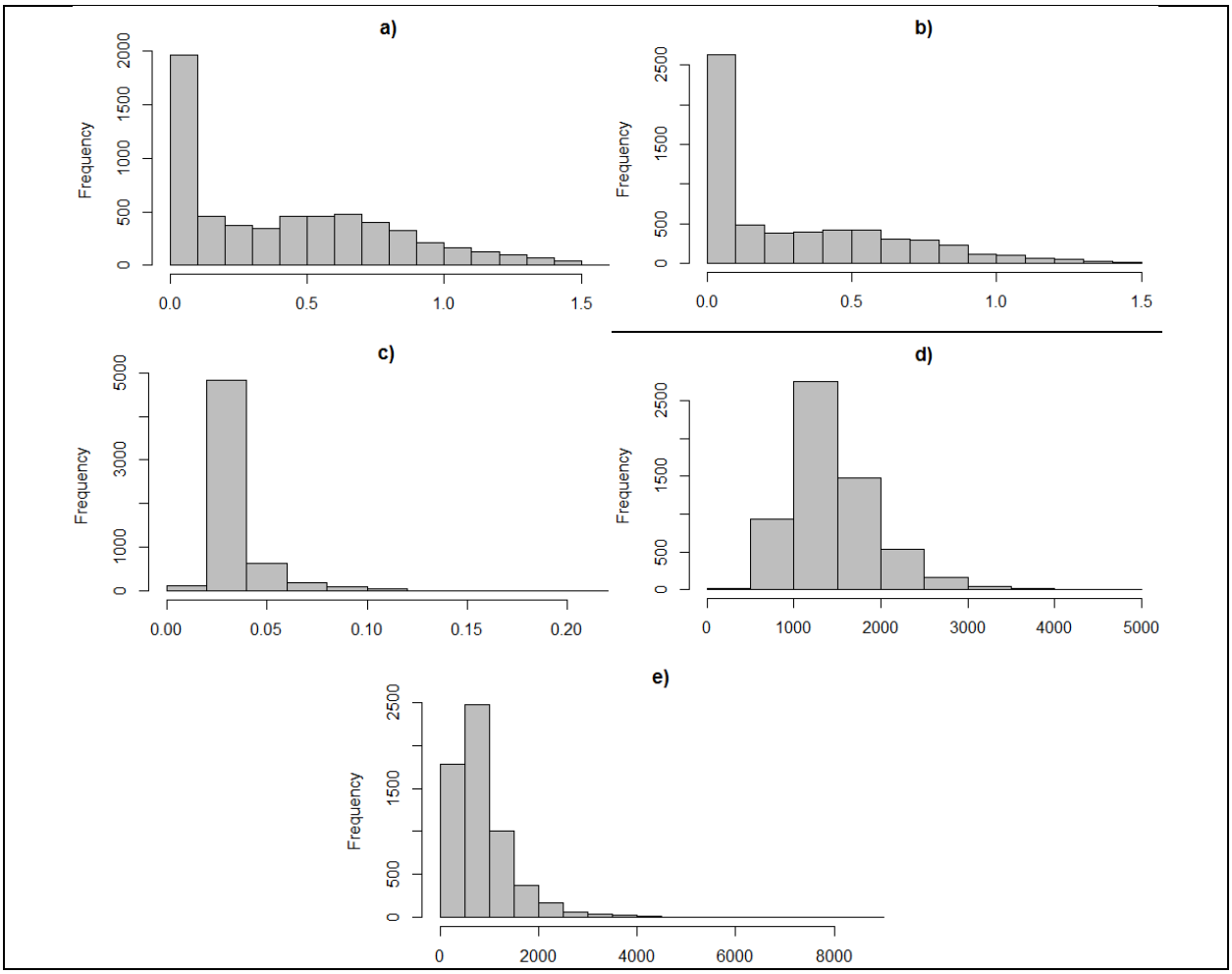

Figure 8.2. Histograms of variables before transformations for x06. a) Med_nsm, b) Min_nsm, c) GLCM_Homog, d) GLCM_Contr, and e) Area.

The presence of zero values were identified for variables where a logarithmic transformation was deemed appropriate. There were zero values found in Med_nsm and Min_nsm which were chosen to be logarithmically transformed per a skewness value of greater than 1 for x01. A constant of 1 was applied to these variables. For x06 there were not variables with a skewness value greater than 1 and a lower limit range of 0. Data distributions after the transformations may be found in Figures 8.3-4.

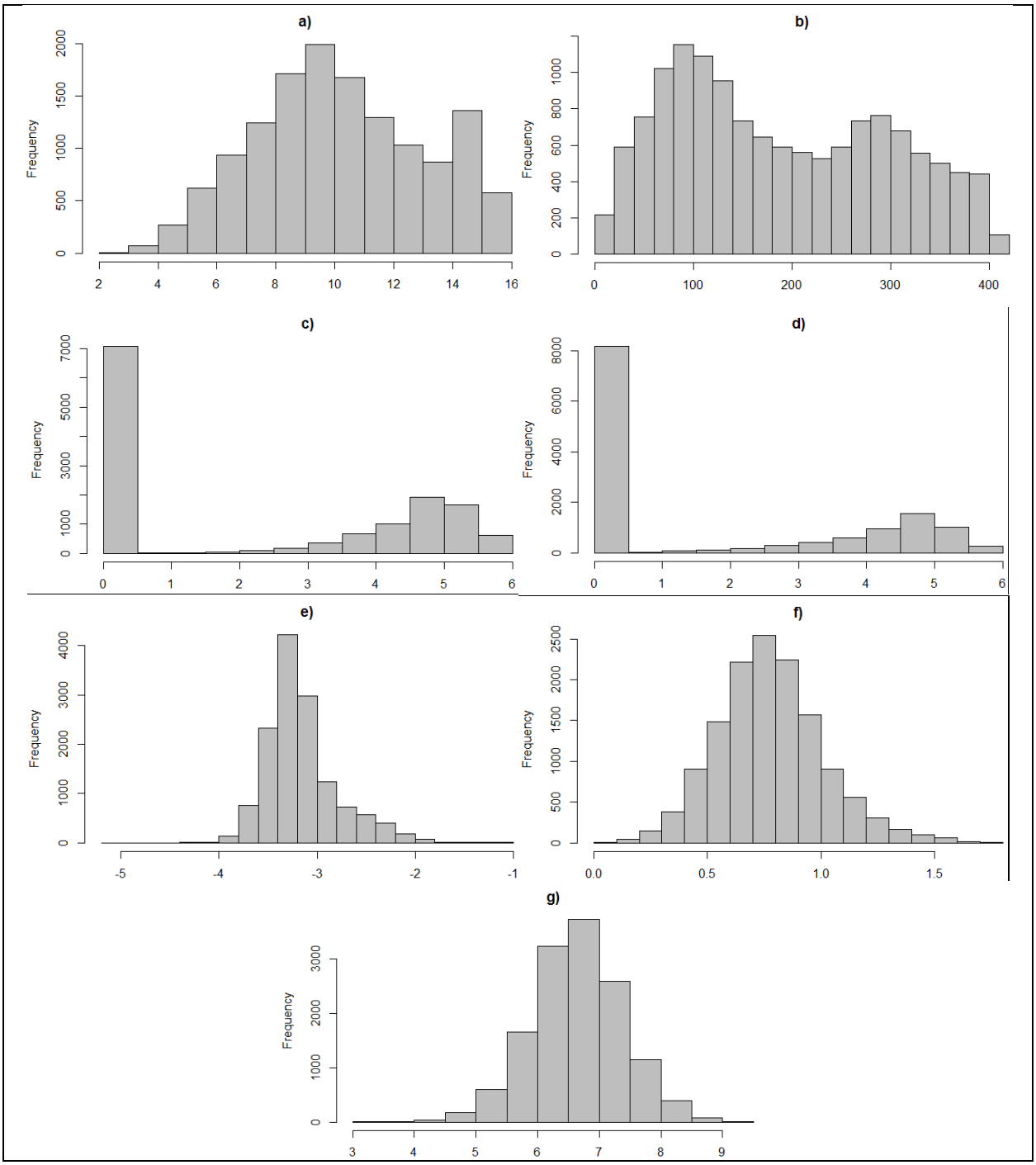

Figure 8.3. Histograms of variables before transformation for x01. a) Mean_red_TranF, b) Mean_gla_TranF, c) Med_nsm_TranF, d) Min_nsm_TranF, e) GLCM_Homog_TranF, f) GLCM_Entro_TranF, and g) Area_TranF.

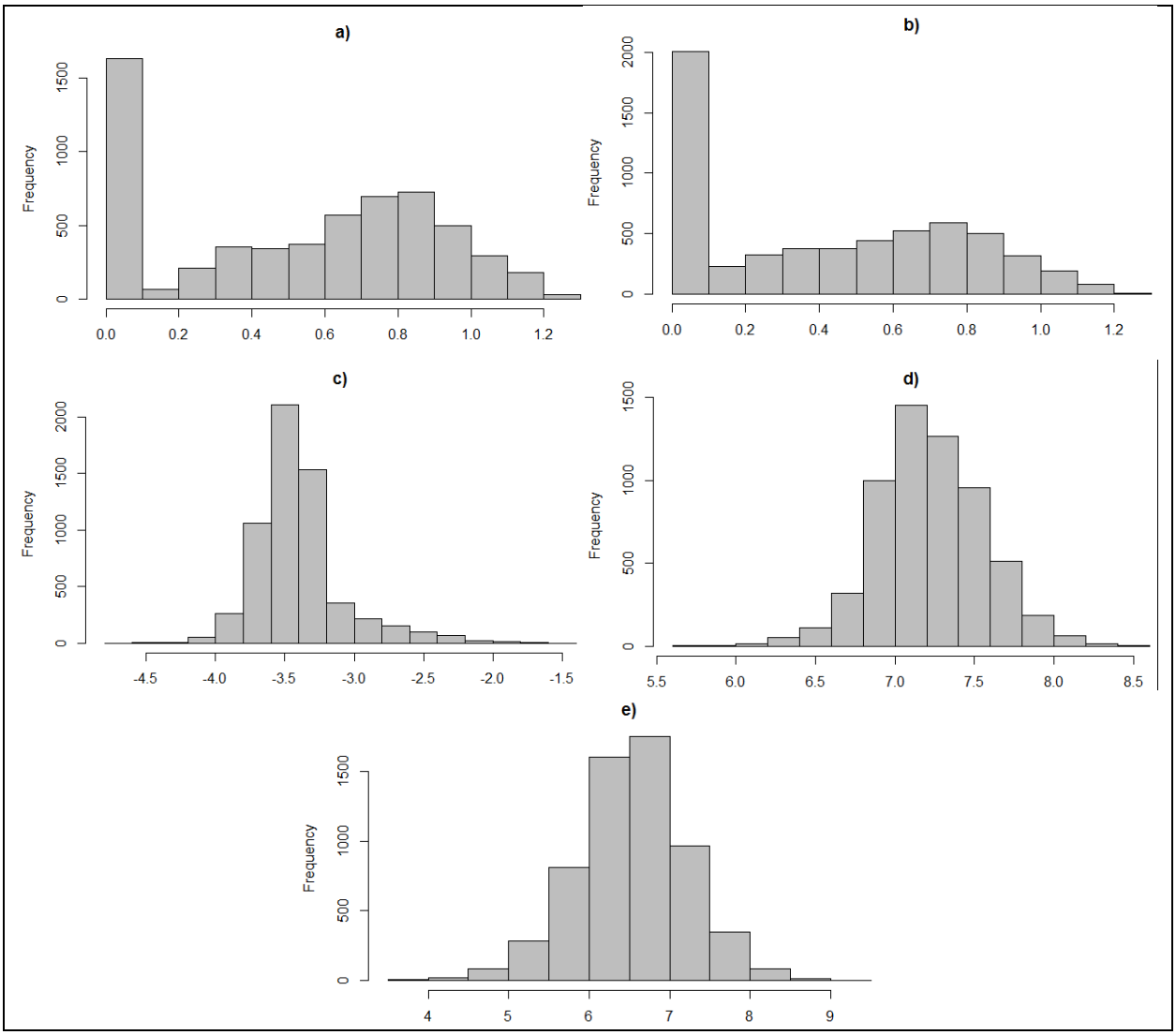

Figure 8.4. Histograms of variables before transformation for x06. a) Med_nsm_TranF, b) Min_nsm_TranF, c) GLCM_Homog_TranF, d) GLCM_Contr_TranF, and e) Area_TranF.

The Mahalanobis distances were calculated for Mean_red_TranF, Mean_gla_TranF, Med_nsm_TranF, Min_nsm_TranF, GLCM_Entro_TranF, and Compactnes x01 variables (Figure 8.5). Mahalanobis distance calculations for x06 were Mean_red, Max_gla, Med_nsm_TranF, Min_nsm_TranF, GLCM_Contr_TranF, Compactnes, and Area_tranF. There were no prominent outliers present for either exclusion plot site that may affect the results.

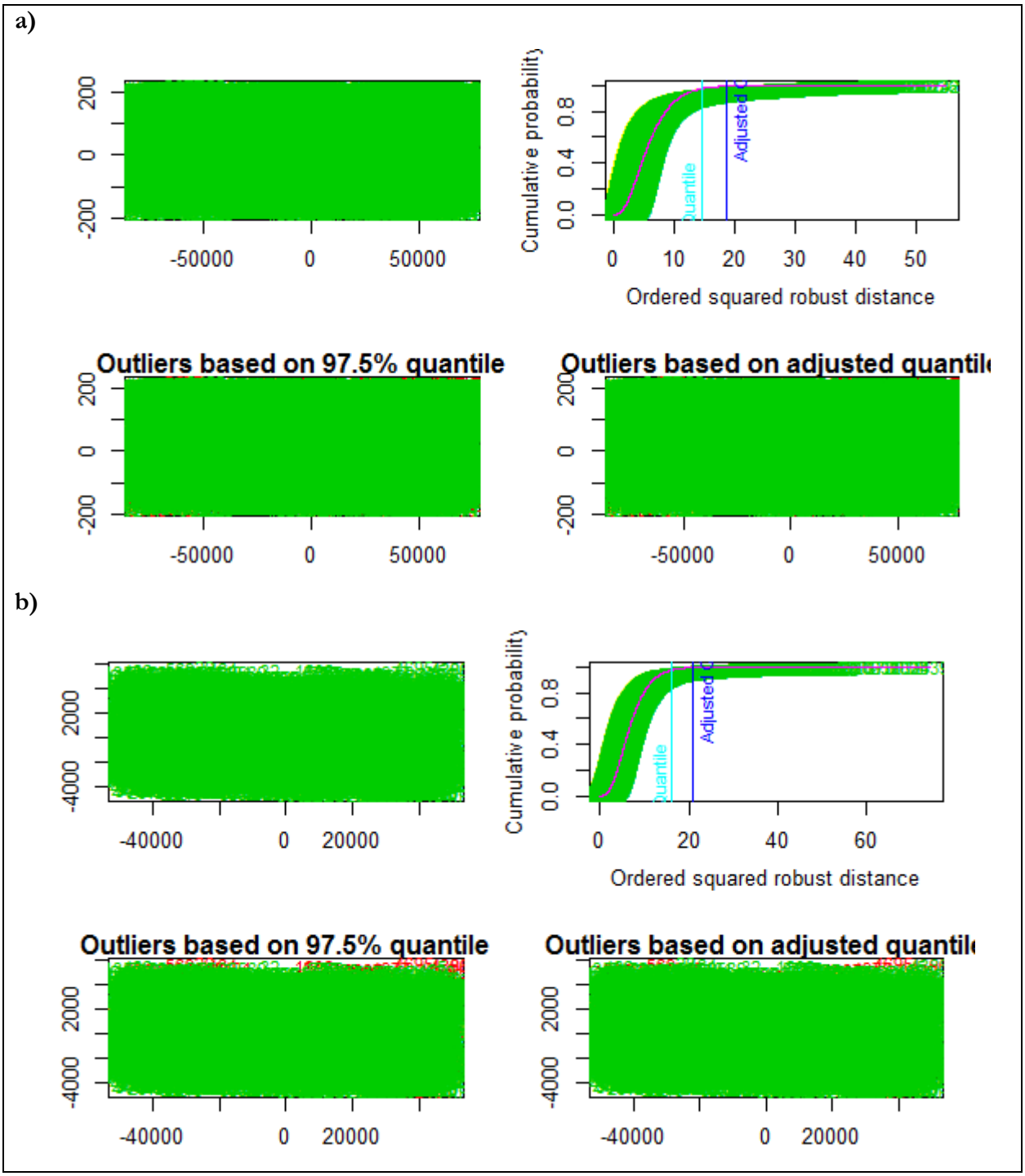

Figure 8.5. Outlier graphs showing the observed versus expected Mahalanobis distances for both exclusion plot sites. a) x01 and b) x06. Top left = distribution of all Mahalanobis distance values. Top right = cumulative probability of outliers with 97.5% quantile and adjusted quantile indicators. Bottom left = outliers shown in red per the 97.5% quantile. Bottom right = outliers shown in red per the adjusted quantile.

Box and whisker plots were recalculated for the transformed Area_TranF and GLCM_Contr_TranF correlated variables found in x06 (Figure 8.6). The spread of values was similar. The GLCM_Contr_TranF ground class had an interquartile range which showed slightly less overlap with plant species than the ground class found in Area_TranF. Furthermore, *Scaevola spinescens* median was in line with the maximum interquartile range value for *Triodia plurinervata* for GLCM_Contr_TranF whilst the median for *Scaevola spinescens* shown in Area_TranF was positioned centrally to the first quartile of *Triodia plurinervata*. The inverse may be found where the GLCM_Contr_TranF median for *Triodia plurinervata* was located at the minimum interquartile range value whilst the median for *Scaevola spinescens* shown in Area_TranF was positioned centrally to the third quartile of *Triodia plurinervata*. *Acacia ligulata* and *Atriplex vesicaria* was relatively similar across both x06 Area_TranF and GLCM_Contr_TranF. GLCM_Contr_TranF showed an increase in dissimilarity between classes when compared with Area_TranF.

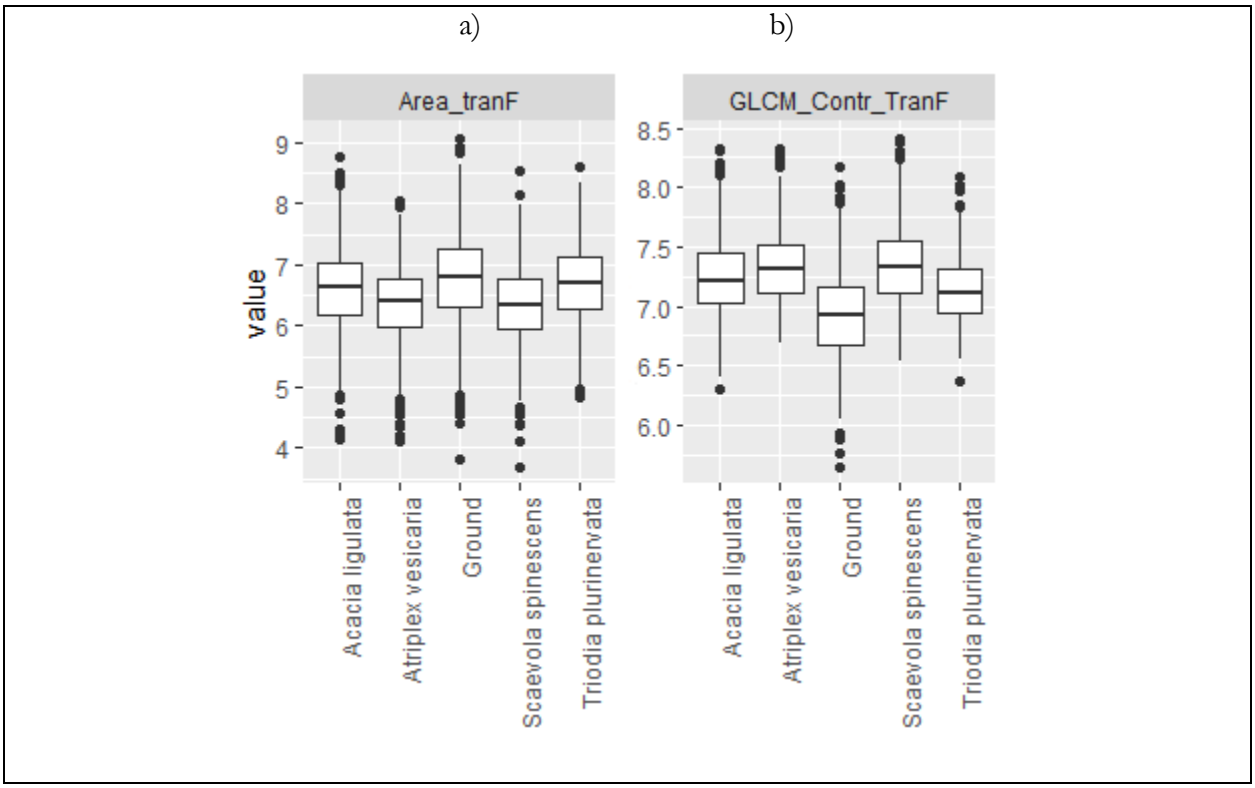

Figure 8.6. Box and whisker plot comparisons showing the spread of transformed variables quantified per species and ground classes for x06. The whiskers encompass samples within 10% to 90% of each distribution with points representing values outside of this range. a) Area_TranF and b) GLCM_Contr_TranF.

8.4. Random Forest scenario testing

The complete set of Kappa and accuracy values for each scenario are shown in Tables 8.4-5.

Table 8.4. Random Forest out of bag Kappa values for 10 iterations per scenario.

Scenario	Kappa values
Exclusion plot site x01	
Scale 20 – all variables	0.6964, 0.6972, 0.6977, 0.6969, 0.6975, 0.6965, 0.6973, 0.6973, 0.6972, 0.6971
Scale 50 – untransformed selected variables	0.5051, 0.5039, 0.5022, 0.506, 0.5034, 0.5027, 0.5037, 0.5058, 0.5033, 0.5017
Scale 50 – transformed selected variables	0.5057, 0.5072, 0.5059, 0.505, 0.5061, 0.5061, 0.5062, 0.5037, 0.51, 0.5075
Scale 50 – all variables	0.6846, 0.6836, 0.6874, 0.6842, 0.6865, 0.6871, 0.6855, 0.6809, 0.6844, 0.6852
Scale 100 – all variables	0.637, 0.6315, 0.6299, 0.6331, 0.6322, 0.6339, 0.6352, 0.6316, 0.6324, 0.634
Exclusion plot site x06	
Scale 20 – all variables	0.6997, 0.6981, 0.6991, 0.6988, 0.6977, 0.7005, 0.6972, 0.6991, 0.6978, 0.6994
Scale 50 – untransformed selected variables	0.5555, 0.559, 0.5579, 0.5556, 0.5609, 0.56, 0.5582, 0.5604, 0.5583, 0.5595
Scale 50 – transformed selected variables	0.5127, 0.5071, 0.5058, 0.5074, 0.5106, 0.5093, 0.5097, 0.5072, 0.5108, 0.5055
Scale 50 – all variables	0.6682, 0.6619, 0.6658, 0.6652, 0.666, 0.6647, 0.6627, 0.6665, 0.6666, 0.6659
Scale 100 – all variables	0.5661, 0.5609, 0.5664, 0.5728, 0.5686, 0.5679, 0.5705, 0.5637, 0.5747, 0.56

Table 8.5. Random Forest out of bag accuracy values for 10 iterations per scenario.

Scenario	Accuracy values
Exclusion plot site x01	
Scale 20 – all variables	0.7508, 0.7515, 0.7518, 0.7512, 0.7517, 0.7509, 0.7516, 0.7515, 0.7515, 0.7514
Scale 50 – untransformed selected variables	0.5895, 0.5884, 0.587, 0.5903, 0.5867, 0.5875, 0.5882, 0.59, 0.5879, 0.5867
Scale 50 – transformed selected variables	0.5899, 0.5912, 0.5901, 0.5894, 0.5901, 0.5904, 0.5903, 0.5882, 0.5934, 0.5914
Scale 50 – all variables	0.7368, 0.7358, 0.7391, 0.7364, 0.7382, 0.7388, 0.7375, 0.7336, 0.7366, 0.7372
Scale 100 – all variables	0.695, 0.6905, 0.6891, 0.6922, 0.691, 0.6925, 0.6935, 0.6905, 0.691, 0.6925
Exclusion plot site x06	
Scale 20 – all variables	0.7785, 0.7774, 0.778, 0.7778, 0.777, 0.779, 0.7766, 0.7779, 0.777, 0.7782
Scale 50 – untransformed selected variables	0.6312, 0.6342, 0.6334, 0.6314, 0.6359, 0.6352, 0.6335, 0.6354, 0.6337, 0.6347
Scale 50 – transformed selected variables	0.6283, 0.6245, 0.6231, 0.6248, 0.6273, 0.6258, 0.6268, 0.6245, 0.6268, 0.6228
Scale 50 – all variables	0.7464, 0.7417, 0.7447, 0.7442, 0.7447, 0.744, 0.7422, 0.7452, 0.7452, 0.7447
Scale 100 – all variables	0.66, 0.6558, 0.6599, 0.6651, 0.6615, 0.661, 0.663, 0.6579, 0.6667, 0.6553