ESTIMATING PALILA (*Loxioides bailleui*) HABITAT USING REMOTE SENSING

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII AT MĀNOA IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

IN

GEOGRAPHY

MAY 2015

By

Jordan Chetcuti

Thesis committee:

Qi Chen, Chairperson
Mathew McGranaghan
Christopher Lepczyk
©Copyright 2015

By

Jordan Chetcuti

ii
ACKNOWLEDGEMENTS

Greg Asner and the Carnegie Airborne Observatory team for flying and providing the LiDAR and Hyperspectral data. Paul Banko, Kevin Brinck and Rick Camp for providing palila count and other station data as well as providing insight into the site and discussion of techniques, including justification for not using occupancy modelling. Paul Banko also took time to show me around the site providing invaluable insight. Qi Chen for making this research possible and for his sharp eyes, help editing and encouragement to aim for excellence. Mathew Mcgranaghan for being an encouraging mentor and a sounding board. Chris Lepeczyk for his patience and quick response every time. Reece Jones for help and encouragement when applying for funds to visit the site.
ABSTRACT

The remaining population of palila (*Loxoioides bailleui*), which is tied closely to mamane (*Sophora chrysophylla*) and naio (*Myoporum sandwicense*) tree species, has been in long-term decline. This decline is caused by black rats (*Ratus ratus*), malaria, predation, and parasitism of moth species that palila consume, and the destruction of the mamane naio woodland by ungulate browsing and invasive flora. Using LiDAR and hyperspectral data collected in 2009, this study derived measures of the mamane and naio habitat. Using Bayesian information criterion (BIC) stepwise selection this study found that LiDAR ground return intensity, area of mamane trees and water band index (WBI) were the best explanatory variables of palila bird count data and that mamane and naio may be used at different scales. The browsing and area of mamane variables support existing knowledge; the canopy water content however suggests a problem for palila in wetter areas, contra to established expectation.
TABLE OF CONTENTS

Acknowledgements .............................................................................................................................................. iii
Abstract ................................................................................................................................................................. iv
Table of Contents .................................................................................................................................................. v
List of tables ......................................................................................................................................................... viii
List of figures ......................................................................................................................................................... ix
Chapter 1. Introduction ........................................................................................................................................ 1
Chapter 2. Research background and questions ............................................................................................. 3
  2.1 Palila .............................................................................................................................................................. 3
    2.1.1 Food sources ............................................................................................................................................... 4
    2.1.2 Ungulate browsing ..................................................................................................................................... 5
    2.1.3 Invertebrate predation and parasitism ..................................................................................................... 6
    2.1.4 Invasive flora ............................................................................................................................................. 7
    2.1.5 Woodland structure .................................................................................................................................... 7
    2.1.6 Predation .................................................................................................................................................... 8
    2.1.7 Disease ..................................................................................................................................................... 9
    2.1.8 Inbreeding and Limited dispersal ability ................................................................................................ 10
    2.1.9 Elevation .................................................................................................................................................. 12
  2.2 Birds, LiDAR and hyperspectral imagery ..................................................................................................... 13
    2.2.1 Crown cover ........................................................................................................................................... 16
    2.2.2 Woodland width ....................................................................................................................................... 17
    2.2.3 Tree age ................................................................................................................................................... 18
    2.2.4 Canopy height and tree height ............................................................................................................... 18
    2.2.5 Total biomass ......................................................................................................................................... 20
    2.2.6 Tree species ............................................................................................................................................ 20
    2.2.7 Understory ............................................................................................................................................. 21
    2.2.8 Elevation ................................................................................................................................................ 22
    2.2.9 Drought .................................................................................................................................................. 23
    2.2.10 Browsing .............................................................................................................................................. 24
    2.2.11 Grain size ............................................................................................................................................. 26
  2.3 Questions .................................................................................................................................................... 26
Chapter 5. Discussion .................................................................................................................................................. 59
Chapter 6. Conclusion ................................................................................................................................................ 70
References ............................................................................................................................................................... 74
Appendix 1 Tree species classification R script ........................................................................................................ 86
Appendix 2 Regression analysis R script .................................................................................................................. 88
Appendix 3 Visual tree comparison .......................................................................................................................... 96
LIST OF TABLES

Table 3.1. Hyperspectral data bands ........................................................................................................31

Table 3.2. Number of points in each original classification of LiDAR .........................................................32

Table 3.3. Showing the variables to be considered in the analysis as well as the code used in R and the results (the codes will have Ma, Na and all at the end, which signify mamane, naio and combined), hypothesised sign of coefficient within the model and the relevance of the variables.....35

Table 4.1. BIC stepwise (k = 4.7) selected negative binominal linear regression models for different scales, and with different starting variables..................................................................................................................55

Table 4.2. Coefficients in single variable models at different radius grain size. ........................................56
LIST OF FIGURES

Figure 1.1. Study site on southwest slopes of Mauna Kea. ................................................................. 2

Figure 3.1. Study site on southwest slopes of Mauna Kea. Area denoted by area containing LiDAR
data and showing bird count stations. ........................................................................................................ 30

Figure 3.2. Stations used in regression analysis relative to extent of all remotely sensed data. ........ 33

Figure 3.3. Palila observed at each station for years of increased survey effort. .................................. 34

Figure 3.4. Shows an example of the smoothing process. a) shows hypothetical CHM with multiple
local maximum. b) shows intermediate smoothed canopy, still showing two local maximum. c)
shows final smoothed CHM, with the focal flow now giving multiple points adjacent to each other.
.................................................................................................................................................................. 38

Figure 3.5. Spectral signature of naio dominated areas in purple (the bottom of the two close lines),
mamane dominated area in green (the top of the two close lines) and an area of conifer plantation in
red (the line most different from the other two). ............................................................................................ 40

Figure 3.6. Showing a histogram of intensity of LiDAR top of canopy returns, showing clearly two
different groupings of intensity...................................................................................................................... 41

Figure 4.1. Comparison of elevation recorded at survey stations and from the DTM ......................... 46

Figure 4.2. Count of trees in the year 2000 survey, against count from the delineation in this study. 47

Figure 4.3. 1600 m² buffer showing number of trees counted in 2000 survey and in the delineation
of the current study, and a true colour image from the hyperspectral data and outlines of delineated
canopies. ......................................................................................................................................................... 48

Figure 4.4. Percentage of mamane from total number of trees in the 2000 survey and the estimation
from the current survey. Black line represents 1:1 line also included count of over, equal and under
estimation and RMSE. .................................................................................................................................... 49
Figure 4.5. LiDAR ground return intensity under mamane and naio tree canopies, outside of protected areas and fence, and in the two protected areas, showing intensity across a range of heights.

Figure 4.6. Mean WBI for mamane trees compared to interpolated inches of rainfall in December 2008 from Frazier et al., (in review.) The black lines are hand drawn showing a line of fit.

Figure 4.7. Mean WBI for naio trees compared to interpolated inches of rainfall in December 2008 and January 2009 from Frazier et al., (in review.) The data has been split into two broad groups, which either fit (blue diamonds) or do not (red circles) fit a relationship of rainfall in the 2008 data. The black lines are hand drawn showing a line of fit.
Chapter 1.
INTRODUCTION

Island bird species are under a higher threat of extinction than in other areas of the world (BirdLife International, 2013a). Many of the species on islands evolved in isolation away from predator, parasites, and pathogens. On the Hawaiian Archipelago, a diverse number of bird species evolved through a process of adaptive radiation from few initial founding individuals. Many of the evolved species are now extinct. One surviving species is the palila (*Loxioides bailleui*). Palila are a species of finch-billed Hawaiian honeycreeper specializing in seed-eating (Banko et al., 2002b), predominately mamane (*Sophora chrysophylla*) seed pods (Banko et al., 2002a).

The population of palila have been reduced to a small area within their former range (van Riper III et al., 1978). Fossil evidence suggests that in addition to being found all over the island of Hawaii, in the past, the range of palila included lowland areas of the island of Oahu (Olson and James, 1982) and Kauai (Burney et al., 2001). The whole species is currently only found on the island of Hawaii, with most of the population restricted to the south and southwest slopes of Mauna Kea, between 2000 and 3000 m (Banko et al., 2002b; Scott et al., 1984), which is likely less than optimal habitat (Scott et al., 1984). Attempts have been made to establish a population on the north slopes of Mauna Kea by translocating birds, although 21% returned to the original population almost immediately (Banko, 2014a). The whole palila population has been in rapid decline since 2003 (BirdLife International, 2013b). Given this rapid decline, the goal of this thesis is to assess which habitat variables best explain palila abundance.

This will be achieved by using remotely sensed LiDAR and hyperspectral data as a tool to quantify habitat variables. These variables will be used to create initial linear models, at different scales, using different tree species and different ways of representing combined tree species. Variables will be removed using stepwise selection. These different models will then be compared to
test hypothesis on scale, tree species and how best to combine different tree species into a single model. Having established which model or models best explain palila count data, the coefficients of each variable will then be interpreted to show if the variable is performing as expected, and what implications this could have for palila.

Figure 1.1. Study site on southwest slopes of Mauna Kea.
Chapter 2.

RESEARCH BACKGROUND AND QUESTIONS

This review is in five parts. The first part of the review will address the literature on palila, looking at previously identified associations between palila and its habitat, existing explanations of the cause of declines within the population and identify gaps in knowledge that this study will answer. The second part of the review will then look for variables that have been measured in other studies quantifying bird habitat and how remote sensing has or has not been, but could be, used in the estimation. This second part will identify how the variables identified in the first part will be measured in this study. The third part will summarise the most pertinent information for this study from with the first two parts. Then in the fourth part, research questions, which have been derived from the literature, will be given, explaining why these are of interest in this study. In the final part, part 5, the hypotheses of this study that will be tested to answer the questions raised will be given.

2.1 Palila

Palila currently live in upland dry forest dominated by mamane (*Sophora chrysophylla*) and naio (*Myoporum sandwicense*) trees above the inversion layer where the climate is cool and dry (Press et al. 2006). Mamane trees are an important food for palila and for their native invertebrate prey (Banko et al., 2002a, 2002b). Palila consume the flowers, buds, young leaves, and immature seeds from within pods. Palila also consume the moth larvae of *Cydia* spp., *Scotorythra* (Geometridae) and *Uresephita polygonalis virescens* (Butler) [Crambidae] (Banko et al., 2002a). *Cydia* spp. the most important invertebrate prey species for palila (Brenner et al., 2002) also feed on the mamane seeds (Banko et al., 2002a). Naio trees area also used by palila as a secondary food source when mamane pods are unavailable.

Many factors are negatively affecting palila. The mamane-naio woodland has experienced long-term degradation due to introduced ungulates (Banko et al., 2013), introduced parasites are
removing the moth the prey of palila (Banko et al., 2002c; Brenner et al., 2002) and invasive flora, Fountain grass (*Pennisetum setaceum*) and German ivy (*Senecio mikanioides*), are displacing palila habitat. The ungulates, parasites and invasive flora are reducing the amount of food available to palila (Hess et al., 1999). Ungulates have also altered the structure of the woodland, multiple elements of which appear to influence palila numbers and their choice of habitat.

Palila, like other island bird species are suffering increased predation by invasive animal species (BirdLife International, 2013a) and suffering from disease spread by invasive mosquitoes (van Riper III et al., 1986). The population viability of palila, may also be at risk due to inbreeding, due to past and future reductions in palila numbers (Fleischer et al., 1994). Palila are a highly site tenacious species, returning to the same areas to breed (Fancy et al., 1993). Palila are unlikely to expand quickly to fill restored habitat. Invasive predators, parasites, flora and mosquitoes, experience an elevation gradient, with less of an effect at higher elevations (Amarasekare, 1994, 1993; Banko et al., 2002c; Brenner et al., 2002). This is likely the reason palila are currently only found at higher elevation (Scott et al., 1984).

### 2.1.1 Food sources

Mamane are essential for palila, as palila consume the immature seeds, flower parts, nectar, buds, and young leaves. Palila consume *Cydia* spp. (Lepidoptera: Tortricidae), which also rely on mamane as a food source (Banko et al., 2002a, 2002b). Palila do consume other plants and invertebrates, for example naio fruit, but only when mamane pods are less available (Banko et al., 2002b). Mamane produce fewer pods and grow less during periods of drought (Banko, 2014b; Banko et al., 2013). Naio react more seriously to drought than mamane, however, in some cases dying (Banko, 2014b; Banko et al., 2013). They are therefore an unreliable fallback food supply when mamane are producing fewer pods due to drought. It is likely more important to look at mamane than naio when modelling palila. Water content should also be looked at due to its effect on the
amount of available food and tree health. Palila are also less associated with mamane saplings preferring more mature trees (Farmer et al., 2014). Tree age may be important when modelling palila.

2.1.2 Ungulate browsing

Historically introduced cattle (*Bos taurus*), goats (*Capra hircus*), and sheep (*Ovis aries*) have, through browsing, destroyed mamane. Feral sheep (*O. aries*), mouflon (*O. gmelini musimon*) and their hybrids, and goats (*C. hircus*) continue to be a problem despite court ordered removal (Hess and Banko, 2011). Long term grazing by ungulates has significantly reduced the historic range of the mamane forest (Banko et al., 2013).

Palila prefer un-browsed native understory (Scott et al., 1984). Ungulates preferentially browse mamane and native understory species over non-native understory species (*Agrostis sanawicerise, Trisetum glomeratum* and *Deschampsia australis*) (Scowcroft and Giffin, 1983). Browsing directly affects palila selection of microhabitat and leads to the expansion of non-native plant species.

Drought causes mamane to produce fewer pods (Lindsey et al., 1997), this is then exacerbated by feral ungulate browsing (Banko et al., 2013). The exacerbation of drought, further supports the need for a measure of water content. Ungulates directly inhibit regeneration of mamane (Scowcroft and Giffin, 1983), but even after removal of ungulates the effects they have had on the balance of species, inhibit regeneration (Hess et al., 1999). For example, additionally non-native grass species inhibit germination and therefore regeneration of mamane (Hess et al., 1999) and mamane regenerate better in mixed stands containing naio, than in stands dominated by mamane alone (Hess et al., 1999), a balance that has been affected by long-term ungulate browsing. Due to lack of association of palila with mamane samplings, a time lag will follow the removal of ungulate browsing which has suppressed mamane regeneration (Brinck and Banko, 2014). Reduced mamane fruit has a
knock on effect on *Cydia* spp. which use mamane pods as larvae (Brenner et al., 2002). *Cydia* spp. are however facing another pressure. A measure of the amount of browsing will likely be a good predictor of palila distribution.

### 2.1.3 Invertebrate predation and parasitism

The moth genus *Cydia* are common worldwide, with several native species in Hawaii (Brenner et al., 2002). A limited number of native parasites of *Cydia* spp. exist in Hawaii, however, new species of parasitic wasps (Hymenoptera) and flies (Diptera) have been introduced (Brenner et al., 2002). Palila also prey on the endemic Hawaiian moth genus *Scotorythra* (Geometridae), which are additionally affected by predatory wasps and ants (Hymenoptera) (Banko et al., 2002c). *Cydia* spp. are found in greater numbers at higher elevation (Brenner et al., 2002), and currently the inverse is true of invasive parasites that affect *Cydia* spp. (Banko et al., 2002c). Parasites, where present, have a significant effect on *Cydia* spp. numbers (Brenner et al., 2002). A limited elevation relationship exists between *Scotorythra* spp. and their introduced predatory invertebrates, with marginally less predation at higher elevation (Banko et al., 2002c). *Scotorythra* spp. are found in lower numbers on western slopes, baring no relationship to distribution of predatory invertebrates (Banko et al., 2002c). Two species of ants extend into palila territory and four species are found below the current elevation of palila, (Banko et al., 2002c). Another palila prey species, the moth *Urosephita polygonalis virescens* (Butler) [Crambidae], is now rarely consumed by palila, likely due to their scarcity, which is potentially due to introduced parasites (Banko et al., 2002c). Given the elevation gradient of these invertebrates, this is possibly another pressure forcing palila to the upper range of its former habitat. It is possible that over time the predatory ant species will continue to move upwards in elevation (Banko et al., 2002c). With the likelihood of temperatures increasing due to global warming, it is also likely that conditions will become more favourable for parasites and predators at higher elevation. Elevation is therefore likely important in predicting the distribution of palila.
2.1.4 Invasive flora

Invasive grasses, as well as being negatively associated with palila microhabitat selection, can contribute to the destruction of the mamane-naio woodland as a whole. Fountain grass (*Pennisetum setaceum*), establishing itself within these woodlands, may act as fuel for fire, devastating these woodlands and allowing for a further increase in fountain grass which is adapted to fire (Scott et al., 1984). Potentially, invasive species can also outcompete native species. In the case of German ivy (*Delairea odorata* (*Senecio mikanioides*)), this species could climb mamane and naio trees themselves competing for light and water and contributing structural stress (Scott et al., 1984), increasing mortality of trees essential for palila. Notably, drought does knock back German ivy (Banko, 2014b). There currently appears to be an elevation gradient in invasive grasses and German ivy, with limited amounts at higher elevation (Amarasekare, 1994; Banko et al., 2002c). This further supports the inclusion of elevation in modelling palila distribution.

2.1.5 Woodland structure

Palila do not occur in all mamane woodlands, likely due to a combination of the factors mentioned so far, but also due to changes in the structure of mamane woodland (Scott et al., 1984). As mentioned previously, native understory is an important component of woodland structure for palila. Branch size is also a factor, with palila unable to perch and feed on smaller branches (van Riper III, 1978). Heavily browsed, or recently regenerated mamane is, therefore, unsuitable for palila (Scott et al., 1984). Scott et al. (1984) measured the width of the woodland along the palila survey transect, across the elevation, but did not include any area of scattered trees, finding a strong correlation of palila numbers with width of woodland. This is consistent with thinking on minimum patch size prevalent within landscape ecology. Large areas of the mamane-naio woodland contain scattered trees, suggesting edge effect is unlikely an issue for palila. It was also found that palila were correlated with “total tree biomass, crown cover and canopy height” (Scott et al. 1984, p. 655).
Mamane-naio tree canopies do not appear to significantly alter the air temperature under the canopy of the trees (Juvik et al., 1993), but do offer shade to palila (Weathers and van Riper III, 1982). Canopies do however significantly attenuate differences in soil temperature (Juvik et al., 1993). The study by Juvik et al. (1993) does not look at the difference in temperature between individual tree canopies and areas of dense closed canopy stands; it may be possible that long-term attenuation is achieved under closed canopy. Palila has an upper critical temperature of 31°C (Psittirostra et al. 1982, p. 667), which suggests that palila can only live at higher elevation (Weathers and van Riper III, 1982). However, the fossil evidence puts palila at low, warm altitudes (Burney et al., 2001; Olson and James, 1982). This suggests that some mechanism existed which enabled palila to escape from higher temperatures likely through migrating farther up and down through elevation following mamane pods and the seasons (Banko et al., 2002c), thereby avoiding very high temperatures. Again, elevation appears to be important when looking at palila distribution, but the woodland width and shade for palila, suggest that patch size should also be considered.

2.1.6 Predation

Hawaiian avifauna evolved with no mammalian predation, with the islands only having two terrestrial mammals, an endemic subspecies, the Hawaiian hoary bat (Lasiurus cinereus semotus) and an endemic species, the Hawaiian monk seal (Monachus schauinslandi) (Olson, 2004). During the various human colonisations of the islands, humans introduced numerous mammal species, either accidentally or intentionally. Among introduced species, there are the black rat (Rattus rattus) (Amarasekare, 1993; Shiels and Drake, 2010) feral cats (Felis catus) and house mice (Mus musculus) (Amarasekare, 1993).

House mice show no evidence of direct predation on palila. Mice do not forage into trees, and their diet is made up of insects and plant matter (Amarasekare, 1994). Mice do not compete for insects, mamane or naio due to the difference in mouse and palila foraging, each at a different strata,
mice in the understory and palila in the over story (Scott et al., 1984). Mice are also susceptible to poisonous alkaloids which mamane seeds contain (Banko et al., 2002a). Feral cats do directly depredate palila (Pletschot and Kelly, 1990), but it is thought the level of predation is very small (Amarasekare, 1994), partially due to the abundance of house mice as prey (Amarasekare, 1993).

Black rats are the most significant predator of palila, accounting for almost all predation at all palila life stages (Amarasekare, 1993). The levels of predation may not be a large driver of palila reduction, however, as rats and palila show different preference for microhabitat (Scott et al., 1984). Palila preferring mamane dominated stands with native understory (Scott et al., 1984) and rats preferring naio (Myoporum sandwicense) dominated stands with invasive grass understory (Amarasekare, 1994; Norman and Baudinette, 1969). Rats do however prefer smaller eggs and choose palila eggs over those of other bird species as the other eggs are too large (Amarasekare, 1994). Rats are limited by the need for additional water to that obtained from their food source (Amarasekare, 1994). Rats exhibit a population gradient with elevation, with lower numbers at higher elevation (Amarasekare, 1993); this may contribute to the restriction of palila to the upper edge of its former range. Rats as a predator for palila further supports the inclusion of elevation in modelling palila, but also suggests water content, although with the opposite association to previously suggested. More water was shown to be important in facilitating more food for palila (Banko, 2014b; Banko et al., 2013). With rats, the relationship is the opposite, with drier areas supporting fewer rats.

2.1.7 Disease

Mosquitoes carry avian malaria which has been shown to be a limiting factor on the population of multiple native bird species in Hawaii (van Riper III et al., 1986). Palila territory is likely limited by mosquito presence (Scott et al., 1984), again forcing palila to the upper edge of its
natural range. Rising temperatures are likely to make conditions more favourable for mosquitoes at higher elevations, further pressing palila to move up slope (Benning et al., 2002).

It has been suggested that some native species may have partial immunity to malaria, and others are beginning to develop some immunity (Scott et al., 2002). Kilpatrick (2006) has suggested that habitat management could lead to palila immunity. Tucker-Mohl et al. (2010) showed that lowland dry forests are not a refuge from avian malaria, due to feral pigs creating mosquito-breeding sites in tree ferns and due to mosquitoes dispersing from nearby residential areas. The upland dry mamane-naio forest is sufficiently far from residential areas, but pools of water could exist or be created that would allow mosquito to breed in an area of the woodland and then disperse across the site.

Rats further stress the palila population, taking palila suffering from malaria that may have survived to pass on resistant genes (Kilpatrick, 2006). Vanderwerf and Smith, (2002) suggested that controlling rats in areas where native Hawaiian birds are killed by Malaria and therefore bird numbers are low, should be given preference over controlling rats in areas where birds are not affected by malaria. This approach may seem counter intuitive, with conservation effort often focused on the bulk of the population, but this approach may result in quicker evolution of resistance (Kilpatrick, 2006). This does suggest, however, that genetic diversity could be lost within palila populations, as only those individuals with immunity survive. Elevation and water are important with the same relationship to mosquitoes to rats, rats also compounding the effect of malaria, which is spread by mosquitoes.

2.1.8 Inbreeding and Limited dispersal ability

As the population of palila has decreased, it is possible that the population could now be experiencing a reduction in fecundity due to inbreeding. Fecundity is affected through “reduced fertility and high incidence of lethal recessives exposed during embryo development” (Pletschet &
Kelly 1990, p. 1019). This is difficult to demonstrate with so many other factors affecting breeding success (Pletschet and Kelly, 1990). Even if the population is not currently showing the effects of inbreeding, further reductions in population could result in an effect. Fleischer et al. (1994) suggest genetic similarity between the main population and the translocated population of palila on Mauna Kea. Neither population were inbred in 1994 (Fleischer et al., 1994). Fleischer et al. (1994) do recognise that their study may not have fully explored all of the possible genetic issues. The sample size was limited due to a fear of harming the population through the sampling technique and only successfully breeding individuals were considered (Fleischer et al., 1994). It is possible that inbred individuals were not breeding successfully. The loci they looked at is subject to more mutation and less variability within the population may exist in genes prone to less mutation (Fleischer et al., 1994). The smaller more isolated of the two populations of palila should have shown lower genetic diversity (Fleischer et al., 1994). This was not seen however, one explanation is that the population was only recently isolated and that the evidence of inbreeding may not have been showing at the time of the study, this seems unlikely however when compared with other bird species that have undergone recent bottlenecks (Fleischer et al., 1994). Another possibility is that palila were moving between the populations, in particular Fleischer et al. (1994) suggest natal dispersal as a possibility. Research into palila dispersal does not support this however.

Limited dispersal ability may not be a factor reducing palila numbers, but is likely to limit recovery. Palila shows strong site tenacity (Fancy et al., 1993), they disperse far less distance than they potentially could. The only factor that seems to cause dispersal is a lack of food (Fancy et al., 1993). Juveniles, usually the highest dispersers in other bird species, were found to breed close to their birth place (Fancy et al., 1993). It may therefore be necessary to translocate individuals to form new site tenacious populations and thereby increase palila numbers, even though 21% returned to the original population in trails between 1997–2006 (Banko, 2014a). It may also be worthwhile
moving individuals between populations. According to the study by Fleischer et al. (1994) this is sound and would not result in a loss of genetic diversity as populations were similar in 1994. The genetics of the palila may benefit from further study, possibly in conjunction with a spatial study. Genetic study is not part of the current study, but the ability to model what is potentially suitable habitat for palila would be useful in comparing where palila are and where they could be (although data covering a larger area would be needed for this future analysis), therefore looking at dispersal ability and possibly habitat connectivity related to genetics.

2.1.9 Elevation

The current elevation limit of palila seems to be due to multiple pressures acting separately across an elevation gradient. Predators of palila have been found less often at higher elevations, thought to be due to the less frequent occurrence of preferred microhabitat (invasive grasses) (Amarasekare, 1994). *Cydia* spp. are also linked to elevation, with more found at high elevation (Brenner et al., 2002), and their invasive parasites and predators found at lower elevations (Banko et al., 2002c). High elevation has currently afforded a refuge from mosquitoes and subsequently malaria that they transmit to avian populations (Benning et al., 2002; Scott et al., 1984). Palila being limited to higher elevations due to temperature is possibly tenuous, due to their former range (Weathers and van Riper III, 1982), however even if this is not one of the issues limiting palila, this may become an issue as global temperatures rise, causing temperature at lower elevations to exceed 31°C all year round. Current elevation limited negative factors are likely to be less limiting in the future, for example invasive invertebrates, which will be able to gradually colonising higher elevations as warming temperatures make higher elevations more appealing. Elevation will likely be important in modelling palila distribution.
2.2 Birds, LiDAR and hyperspectral imagery.

Palila are not alone, bird species are under threat worldwide, in the biggest modern threat to biodiversity. Understanding the associations of bird species to their habitats are essential in ensuring that conservation effort is focusing on the true causes of decline and allowing efforts to focus on ameliorating deleterious effects on bird numbers.

Remotely sensed data allows for the continuous assessment of habitat across space, which allows for better quantification of the vegetation structures influencing bird abundance. This has led to multiple studies using discrete return Light Detection And Ranging (LiDAR) to look at bird diversity. LiDAR has been used in North American woodland (Jones et al., 2013; Lesak et al., 2011; Seavy et al., 2009) and looking at the numbers of bird in individual bird species in North America (Garcia-Feced et al., 2011; Jianchu et al., 2005; Martinuzzi et al., 2009; Smart et al., 2012), Britain (Bellamy et al., 2009; Bradbury et al., 2005; Hinsley et al., 2009) and central Europe (Müller et al., 2009a; Zellweger et al., 2013). In their North American study, Jones et al. (2013) combined LiDAR with hyperspectral imagery to add tree classification to the LiDAR derived vegetation structure in an assessment of bird richness. A similar technique combining LiDAR and hyperspectral data may be useful in the case of palila.

Most studies of the usefulness of LiDAR in modelling either bird species diversity or looking at individual bird species have been conducted in temperate climates in North America or Western Europe. Two studies have used LiDAR to look at avian abundance in Hawaii. Boelman et al. (2007) used LiDAR in combination with hyperspectral aerial imagery to compare a native and non-native transect of vegetation and the effect of invasive fauna on avian abundance. Flaspohler et al. (2010) used LiDAR in a study of woodland fragment size to measure vegetation volume, maximum tree height and canopy height heterogeneity to predict bird richness. Both studies contained elements indicating the usefulness of some of the variables they derived from LiDAR in modelling bird
abundance or bird diversity. Neither study however exhaustively investigated the variables that are derivable from LiDAR nor did they apply their research directly to a single avian species with its particular requirements. Davies and Asner, (2014) used the Flasphohler et al. (2010) paper to make generalisations about Hawaiian birds, however, Hawaii is a diverse environment with many different bird species occupying different niches. These two studies cannot be generalised for all Hawaiian bird species and there is still value in conducting research into different bird species in the different environments of Hawaii.

Studies of the association of birds with habitat usually look at either one or a few different measures of birds (presence-absence, occupancy, abundance, or diversity). Any variable affecting one is likely to affect the others, for example, if tree height influences presence-absence, then it will also affect abundance, and given the fact that different species prefer different heights of tree, diversity.

Traditionally, quantification of the variables affecting birds used field measurement. Field measurements are time consuming and therefore costly, leading to constraints in the areas that are covered and sampling of the landscape. Remote sensing allows for continuous measurement across the landscape and in the case of LiDAR vertically within vegetation. Remote sensing is the measurement of variables from afar, either passively through the recording of reflected or emitted electromagnetic radiation, an example of which is hyperspectral imaging or actively through the active emission of radiation by the measurement device and the recording of the backscatter. Two examples of which are Radar and LiDAR. LiDAR (Light Detection And Ranging) is the emission of electromagnetic radiation, often infra-red radiation and the recording of the backscattered energy. Laser energy is emitted as a pulse of radiation and the backscatter recorded either discretely (discrete return LiDAR) or continuously (full waveform LiDAR). Discrete return LiDAR usually records between 2 – 5 distinct returns coinciding with peaks in the amplitude of backscattered returns.
(Jensen, 2007). Hyperspectral imagery differs from multispectral imagery having higher spectral resolution and can be obtained either from satellite or airborne. LiDAR can obtain measurements from three platforms, satellite, airborne or ground-based and gives distance from the platform to the reflecting surface as well as information on the intensity of the return. Most frequently, the first returns and the last returns are each interpolated to give the height of the uppermost surface and ground surface, however, LiDARs ability to penetrate through the gaps in vegetation and the difference in the intensity of the returns, can be used to infer a considerable amount of information beyond the elevation and upper surface of vegetation.

This second part of the review will focus mainly on airborne discrete return LiDAR, as this is available for the study, and identify variables affecting palila numbers that can and cannot be measured using airborne this data. Where LiDAR is ill suited to or inferior, hyperspectral imagery will be discussed.

This review has so far shown that predation from rats, parasitism and predation of invertebrates on palila food sources, drought, invasive flora, ungulate browsing, and mosquito born avian malaria, all negatively impact palila numbers. Many of these variables have a relationship with elevation. Palila are dependent on mamane and naio for food and palila numbers have previously been linked to total biomass, crown cover, canopy height, age, woodland width, and branch size. The second part of this review will focus only on these variables, showing similarity and difference with other studied bird species and if these variables can be measured using LiDAR or hyperspectral data. Not all variables affecting an animal population are measurable using remote sensing. These latent variables can only be inferred through other variables. The presence or densities of rat population are latent as are parasitism and predation of invertebrates and malaria. Each of these latent variables, however, relates to elevation or available water.
Woodland is an extensively studied land cover type with many studies examining avian use of woodland. Woodlands are three-dimensional, containing different plant species at different strata; consequently, woodlands are dividable into a huge array of different factors affecting birds. Woodlands can be looked at in terms of vertical structure (Blanco and Velasco, 1996); the canopy, understory, field layer and ground layer. The canopy layer consists of the larger mature trees. Understory is the layer of immature canopy trees and shrub species growing in the reduced light under the canopy. The field layer, also called the herb layer, is a layer of herbs, ferns, and grasses growing on the forest floor. The ground layer are mosses, likens and fungi growing below the field layer. Mamane-naio woodland, only really has a canopy layer and field or herb layer. No real understory exists.

2.2.1 Crown cover

The terms crown cover, canopy cover, canopy closure and canopy density, are terms frequently used interchangeably. Crown cover as described by Scott et al. (1984) in relation to palila, is how closed the woodland is, given as a percentage of the ground the canopy is covering. This definition seems the same as canopy cover. Canopy cover is a measure of tree or tree species dominance over a site (Jennings et al., 1999), usually expressed in terms of percentage cover or as open, scattered and closed canopy. Canopy closure “is the proportion of the sky hemisphere obscured by vegetation when viewed from a single point” (Jennings et al. 1999, p. 62). It gives a better measurement of the light regime and microclimate of a particular location (Jennings et al., 1999). Jennings et al. (1999) states that canopy density is a synonym for canopy closure. Anecdotally there does appear to be some sort of difference, canopy density appearing to include three-dimensional measures of canopy. In a poster for the 2010 National Forest Health Monitoring poster presentation Randolph & Stolte (2010) state that canopy density is related, but different to canopy cover and canopy closure.
Canopy cover is a significant indicator of high abundance of bird species diversity (Sallabanks et al., 2006a; White et al., 2013), although the presence of a high cover of particular tree species can also lead to a reduction in species abundance (Kirk and Hobson, 2001; Matlock and Edwards, 2006; White et al., 2013). Canopy closure levels are important for different species, with some species preferring low closure and other intermediate or high (Kirk and Hobson, 2001; McWethy et al., 2009).

Alexander et al. (2013) demonstrated a technique for converting the point cloud from airborne discrete return LiDAR into canopy cover and closure and their relationship to light levels. Canopy cover was calculated by creating polygons representing each point return from the canopy, more than 1 m above the ground, canopy cover was then calculated as the percentage coverage of the whole. Smith et al. (2009) simply calculated canopy cover as the returns above a particular height above the ground divided by the total number of returns. Due to the subjective nature of using a densitometer Smith et al. (2009) suggest that LiDAR derived canopy cover provides more accuracy than does field measurement. The canopy closure is more highly correlated with measured light levels than canopy cover and this study will not use canopy closure, as palila are not ground birds. Canopy cover is however related to total area of the trees, especially when using a defined area to amalgamate area. The 1 m above ground cut off will be useful in defining what area to include when looking at area for palila.

2.2.2 Woodland width

Width is likely a measure of patch size for palila. Mamane-naio woodland is scattered and therefore a detectable edge effect is unlikely. Searching for woodland width in relation to avian or bird habitat only returns one relevant paper (Groom and Grubb, 2002), states that “Forest area is known to be correlated with bird-species diversity. Most researchers of riparian forest avifauna have used riparian woodland width as an index of forest area.” Patch size is an important factor in bird
species diversity with certain species needing large patches to survive. Beier et al. (2002) showed that
the number of species found within larger patches of woodland was substantially higher per hectare
than that of smaller patches; further, smaller patches did not show higher diversity in close proximity
to large patches. The clumps of trees with continuous canopy may act as patches. LiDAR can be
used to give the patch size by measuring the mean area of these contiguous areas. This will be used
as patch size in place of the overall width of the woodland.

2.2.3 Tree age

Different ages of trees within a woodland as well as the mixes of different ages are important
for different bird species (Hewson et al., 2011; Kirk and Hobson, 2001; Poulin et al., 2008;
Sallabanks et al., 2006b; Wunderle Jr., 1999). LiDAR cannot be used in the direct ageing of trees;
rather the age is determined by allometric equations from the size of the trees. Fan et al. (2003)
stated that basal area is strongly correlated with stand age. A further section will discuss using
LiDAR to determine the size of trees. Tree height is itself related to age (Edwards and Mason,
2006). A relationship between age and height is implied in the study by Brinck & Banko (2014)
looking at mamane and naio woodland. “Height of a tree can often be treated as an indicator of tree
maturity…” (Brinck & Banko 2014, p. 5), but the study does not contain any analysis comparing the
age of the trees to the height.

2.2.4 Canopy height and tree height

Canopy height is one of the best indicators of bird species richness and abundance (Kirk and
Hobson, 2001; Matlock and Edwards, 2006). The definition “Canopy height (or stand height), h (m),
of a forest should represent the “average” height of the tallest tree stratum in a stand” (Nakai et al.
2010, p. 1225) makes it clear that a measure of individual tree height although included in a mean
canopy height, provides different information. Tree height allows for finer scale analysis and the
inclusion of information that would be lost if only mean canopy height is considered.
Canopy height can be calculated from LiDAR surface returns, the digital surface model (DSM), minus the interpolated LiDAR ground returns, DTM. This calculation may lead to an underestimate of mean tree height of a few metres (Nilsson, 1996) due to the vertical accuracy of the LiDAR system being used (15 cm) and errors in the DSM (Popescu et al., 2002). Within a single study this error is likely not an issue when considering the tree height as a variable for avian habitat modelling. The height across the site is relative, with certain areas having taller trees than others area do, but the actual height being less important.

To calculate individual tree height, individual trees need to be identified. LiDAR can be used to detect individual tree location as well as the extent of individual canopies (Chen et al., 2006; Hu et al., 2014; Lee and Lucas, 2007; McCombs et al., 2003). The identification of individual trees, suffers from errors of omission and commission. Small trees or trees with merged canopies may appear as one tree, or multiple local maximum within a single canopy may be counted as multiple trees (Chen et al., 2006). It is easier to identify individual trees if the shape of the individual crowns is simple, for example that of a close grown conifer (McCombs et al., 2003). With shapes that are more complex it is necessary to attempt to filter the local maximum using the height of the maximum and relating this to canopy size or through some sort of smoothing to remove additional local maximum. Watershed segmentation is then used to separate crowns (Chen et al., 2006; Hu et al., 2014). LiDAR is superior to hyperspectral imagery in identifying individual crowns (Dalponte et al., 2014).

Having identified the trees, the height of each individual tree can be derived (Kwak et al., 2007). These derived heights are still however subject to underestimation. The tops of the trees may be missed or not be large enough to cause a LiDAR return. It may therefore be necessary to estimate an addition to the heights measured (Wang and Glenn, 2008). Although Chen (2007) argues that LiDAR provides a more accurate height measurement than field methods, against which the LiDAR measurements are compared. Tree height will be used as an approximation of age in this study.
Canopy height is not of interest on its own when looking at palila, but will be used in the calculation of volume under the canopy.

### 2.2.5 Total biomass

There are different measures of tree size. In traditional field measurement, it is convenient to measure diameter at breast height (DBH) to give an estimate of the size of the tree. Basal area is a summation of the area of trunk calculated from DBH. Due to the history of field measurement the literature shows a relationship of bird abundance to both DBH (Müller et al., 2009b; Nur et al., 2008; Poulin et al., 2008; White et al., 2013) and basal area (Hanowski et al., 2005). Basal area is also a good proxy for stand age and subsequently the abundance of trees containing cavities which certain birds use for nesting (Fan et al., 2003). There is no particular need however to convert to these early stages of size estimation, i.e. basal area or DBH, for use in modelling bird abundance, tree volume would be sufficient (Caston and Samantha, 2014; Clark et al., 1983; Sanitjan and Chen, 2009).

Within even aged stands basal area and DBH can be estimated by using the dominant (maximum) tree height and allometric equations (Tesfamichael et al., 2010). Having identified individual crowns, the volume of the tree can be calculated as the volume under the canopy height model derived from the LiDAR data (Chen et al., 2007, 2006). Volume will be relevant to palila both in terms of the total amount of trees that are available to produce food, but will also likely be related to age of the trees.

### 2.2.6 Tree species

Tree species is important in determining species abundance due to association of a particular bird with a particular tree species (Drever and Martin, 2010; Klaus and Keyes, 2007; Lee and Marsden, 2008; Müller et al., 2009a; Whelan, 2001) or through abundance through increased heterogeneity of species (McD. Sweeney et al., 2010). Too high a proportion of an individual species
within a canopy is negatively associated with bird species diversity (Kirk and Hobson, 2001). Different assemblages of trees leading to different assemblages of birds (Lee and Rotenberry, 2005) more species of tree with more species of bird (Fuller and Rothery, 2013; Matlock and Edwards, 2006).

Without identifying individual tree locations, LiDAR can be used in conjunction with imagery to classify the woodland into vegetation classes of similar species composition (Hill and Thomson, 2005). Different mixes of species may also lead to different proportions of returns, allowing approximate proportions of individual tree species to be determined (Moffiet et al., 2005). Having identified individual canopies, structural variables can be used to differentiate individual species (Zhang and Liu, 2013).

Individual tree species can be differentiated using multiple LiDAR derived variables potentially without the need for multispectral or hyperspectral imagery. Holmgren et al. (2008) used shape of the crown, tree height, and tree base height, but also used intensity of returns. Korpela et al. (2010) also showed that LiDAR return intensity was different between trees species, although depending on the growing conditions more similar between similar conifers or deciduous tree species, and between living and dead trees. Moffiet et al. (2005) also conclude that average LiDAR return intensity and variation may allow for differentiation of species as long as the LiDAR pulse is relatively consistent over the whole site. These three studies by Holmgren et al. (2008), Korpela et al. (2010) and Moffiet et al. (2005) suggest that under the right circumstances LiDAR intensity could be used for classification. Given that only two tree species, mamane and naio, dominate the forest, LiDAR intensity will be used in the classification of the species in this study.

**2.2.7 Understory**

Understory is the layer of immature canopy trees and shrub species growing in the reduced light under the canopy. Sometimes the field layer, also called the herb layer, a layer of herbs, ferns
and grasses growing on the forest floor, is also included within the understory. Not all woodlands contain an understory, but where present the bird community is more diverse (Heyman, 2010). In recording understory returns from LiDAR it is important to differentiate between overstory and understory. This can be accomplished by using an arbitrary height thresholds, for example, a one meter cut off for the height of the ground vegetation. This cut off can be decided using prior knowledge of the base of different tree species (Hill and Broughton, 2009), a derived cut off from sampling the vegetation in the field (Wing et al., 2012), or through derivation from the LiDAR data, for example identifying the base of the overstory trees (Korpela et al., 2012). Grasses would be part of the field layer and despite Korpela (2008) having some limited success in differentiating lichen mats on forest floor, it is unlikely that LiDAR could be used to differentiate between native and non-native field layer in mamane-naio forest.

2.2.8 Elevation

In the case of palila, elevation has been shown to influence multiple actors, which then act on palila. Other studies do not go into specifics, but elevation has frequently been shown to be associated with bird numbers (Dickson et al., 2009; Pedersen et al., 2013; Wen-bo et al., 2007; Wunderle Jr., 1999). Dickson et al. (2009) looked at three species and found elevation to be either, positively or negatively associated with each. Wunderle Jr. (1999) found more species at lower elevations. Even if a species numbers are not correlated with elevation, if the species is limited to a narrow range of elevations then elevation is still important (Wen-bo et al., 2007). A digital terrain model (DTM) can be derived from last discrete LiDAR returns which are interpolated to create a continuous surface (Lloyd and Atkinson, 2006). Limitations do exist in the creation of a DTM, particularly under high vegetation, where penetration may be limited. Coveney (2013) for example found that, error in point elevation stated by LiDAR data providers, was for paved surfaces and that the error on vegetated areas was higher. They found that within vegetated land covers the error was
highest in “deep canopies (forest), deep/dense canopies (brush/scrub) and shallow/dense canopies (open areas)” (Coveney 2013, p. 480). Reutebuch & McGaughy (2003) found that using high resolution LiDAR that, although a relationship did exist between canopy density and DTM error, that difference different densities was very small, and that high resolution LiDAR provided a very accurate DTM.

2.2.9 Drought

Jones et al. (2013) looked at groups of bird species and found that vegetation water content appeared to have the largest impact on differences in bird diversity both in open and in forest habitat, this implies that a large number of bird species require vegetation containing large amounts of water. More forest bird species were associated with higher water content (Jones et al., 2013). Diversity of differing amounts of water is also likely important, with different species adapted to different niches. It is therefore surprising how infrequently water is included as a variable outside of wetland or marine bird modelling.

Water content is calculated by looking at the ratio between water absorption wavelengths. LiDAR uses only one wavelength so is unusable in calculating water content. Jones et al. (2013) used hyperspectral data to establish foliar water content in their study of bird diversity. Other than this use by Jones et al. (2013) canopy water content has not been used in avian modelling. Foliar or canopy water content has been calculated from hyperspectral data for other uses, for example, in classifying between native and not native flora (Asner et al., 2009). Many studies use high wavelength water absorption bands (Asner et al., 2009; Casas et al., 2014; Fensholt and Sandholt, 2003; Gao and Goetz, 1990; Qi et al., 2014; Yi et al., 2014). Multiple measures of water stress and drought, use absorption bands at 819, 857, 900, 970, 1241, 1599, 1640, 1649 and 2130 nm (EXELIS visual information solutions, 2014). The data available for the palila site has a maximum wavelength
of 1050 nm. Clevers et al. (2010) used the lower water absorption band at 970 nm. The water band index (WBI) uses the ratio of 900 and 970 nm.

WBI has been used previously, to quantify site quality, defined as total available water (TAW) (Cho et al., 2010). Nagy and Tamás (2013) used WBI to assess water stress in apple and pear orchards. They however found that WBI was more accurate when adjusting WBI to use 886 and 937 nm for these species. This does run the risk of interference from the 940 nm water vapour absorption band (Gao and Goetz, 1990). Claudio et al. (2006) used WBI to look at water status and evapotranspiration along transects in a semiarid environment, but did not look at individual trees even though the resolution was high enough that species were comparable. This study also used a ground-based system, not aerial or satellite. It was shown that different species had different WBI values and that they differed in the strength of their response to drought (Claudio et al., 2006). Claudio et al. (2006) also showed that WBI described the rainfall well. WBI will be used to measure the water content in the mamane and naio trees.

2.2.10 Browsing

Browsing has been shown in many studies to negatively affect bird habitat quality, affecting bird numbers and diversity (Allombert et al., 2005; Graham et al., 2014; Holt et al., 2014, 2013; Pedersen et al., 2007; Rae et al., 2014). Rae et al. (2014) did show that early successional bird species were positively associated with the browsing of moose; forest interior and generalist bird species were negatively affected. This was also the case for the other studies. Most of these studies compared exclusion (Holt et al., 2014, 2013; Pedersen et al., 2007) or islands with no deer (Allombert et al., 2005) with the local status quo, while Rae et al. (2014) compared higher browsing than the status quo. This background browsing in all of these cases is very high due to the removal of natural predators that control browser numbers. It is likely that avian diversity would be higher in areas with low levels of browsing than in areas with no browsing due to increased heterogeneity.
Even in the study by Berger et al. (2001) which focuses on the affect that the removal of predators has had on avian diversity they only examined the landscape with no predators.

Given the importance of browsing to diversity, it is surprising that few studies have attempted to use LiDAR to model the amount of browsing pressure. Only three studies (Coops et al., 2010; Levick et al., 2010; Lone et al., 2014) used LiDAR in relation to browsers, only one looking at the damage caused by browsers. Coops et al. (2010) was looking to establish the quality of deer habitat for mule deer, which are in decline, by looking at the elevation, canopy cover, and solar radiation. Lone et al. (2014) used LiDAR to measure the amount of biomass available for moose using full waveform LiDAR to calculate canopy cover and understory density. Canopy density is mentioned in relation to canopy cover, but again this study is not looking at browsing (Lone et al., 2014). Levick et al. (2010) did compare protected or browsed termite mounds, using canopy height and cover. This appears to be a relatively simple landscape with the termite mound being either vegetated or not.

It is possible that a metric could look more directly at browsing. Hauglin et al. (2013) for example used LiDAR to look at branch biomass with the aim of quantifying the quantity of logging residue that would remain after felling. Korpela et al. (2013) looked at the relationship of LiDAR return intensity in relationship to the silhouette of tree branches. Korpela et al. (2013) point out that a laser pulse intercepting a branch (branch and foliage) will have less energy when reaching the ground, therefore based on the intensity of the ground return the size of the branches intercepted are quantified. Browsers consume biomass and reduce branch size. This suggests that branch size and browsing can be looked at using LiDAR intensity. Mean LiDAR ground return intensity will be used in this study in an attempt to measure levels of browsing.
2.2.11 Grain size

Meyer and Thuiller, (2006) found that species distribution is easier to predict when a grain size similar to the home range of the species is used in the aggregation of variables, although they did suggest that multigrain multiple hierarchical equations predict species distribution better than single grain, this appeared to be over many habitat patches within a fragmented landscape. Seavy et al. (2009) conducted a multi-scale analysis in an attempt to determine and match the spatial scale of the bird species they were modelling. Fancy et al. (1993) determined the spatial scale that palila use the mamane-naio woodland, with a median distance between observation locations and the centre of bird activity ranging between 388 m in 1991 and 641 m in 1988. Fancy et al. (1993) found no significant difference between palila home ranges of 2.66 km$^2$ and 3.70 km$^2$. The standard errors of these measurements were high, ranging from one to 196 m on the median distances and 0.40 and 1.46 km$^2$ on the home range size (Fancy et al. 1993). Farmer et al. (2014) estimated daytime kernel home range from radio tagged palila. The mean of all overlapping group means was 10.25 ha $\pm$ 0.36(SE) (calculated from Farmer et al. (2014)).

Distances between daytime locations and night-time roosts varied considerably (~0-5 km) in the study by Farmer et al. (2014). Farmer et al. (2014) points out that habitat selection between daytime and night time differ, with palila showing a preference for mixed mamane and naio stands, but does not give specific enough information to model proximity to prime roosting locations. This study will therefore not attempt to model proximity to night roosting habitat although this will likely have some effect on bird numbers at particular locations during daytime.

2.3 Questions

Through a review of the literature in the previous sections, certain variables stand out as being important in explaining palila distribution and their decline. Mamane produce fewer pods and grow less, and naio begin to die when deprived of water (Banko, 2014b; Banko et al., 2013).
Elevation has been shown to be associated with predators of palila (Amarasekare, 1993), mosquitoes and malaria (Scott et al., 1984), parasites of their invertebrate prey (Banko et al., 2002c) and invasive flora (Amarasekare, 1994; Banko et al., 2002c), with all being more prevalent at lower elevation. Some of the invertebrate prey of palila are also found in higher numbers at higher elevation (Brenner et al., 2002). It has also been suggested that palila are limited by temperature and therefore elevation, unable to tolerate temperatures over 31°C, although they were found at lower elevation in the past (Weathers and van Riper III, 1982). Palila are dependent on mamane and naio for their food (Banko et al., 2002a, 2002b), therefore the amount of available biomass, measured as either the volume or area of trees should dictate where palila are found. Ungulates have been shown to reduce the extent of mamane naio woodland (Banko et al., 2013) and to reduce available biomass (Scott et al., 1984), therefore, a measure of ungulate browsing should predict palila numbers. Palila have also been shown to prefer more mature trees, anecdotally linked to tree height (Brinck and Banko, 2014). Patch size has not previously been shown to be related to palila numbers, however total woodland width was shown to be linked by (Scott et al., 1984). Naio are considered a backup food source for palila with palila preferring mamane and associated invertebrates (Banko et al., 2002a; Brenner et al., 2002). In measuring variables and then relating them to palila count data collected at stations, it is necessary to aggregate the variables to an appropriate grain size. Meyer and Thuiller, (2006) suggest that when modeling species habitat association the grain size should be that of the home range of the species in question.

This study seeks to answer these three questions, 1) which variables are most significant at explaining the 2009 palila count, 2) are mamane variables more useful than naio variables at explaining the 2009 palila count and 3) does grain size matter when modelling palila and if so what is the optimal grain size.
2.4 Hypotheses

1. The model that best explains the 2009 palila count, will include the following variables:
   a. WBI
   b. Total volume
   c. Mean LiDAR ground return intensity
   d. Elevation
   e. Mean contiguous canopy.

   Area covered by trees and tree height will be unimportant predictors as both will be covered by total volume.

2. With the exception of Mean LiDAR ground return intensity, all the coefficients are expected to be positive.

3. Mamane variables will fit the 2009 palila count better than naio variables.

4. Combined mamane and naio variables will explain the palila count as well as using mamane and naio variables separately.

5. The grain size equal in size to the home range of palila, will better explain the 2009 palila count, than other grain sizes.
Chapter 3.

METHODOLOGY

3.1 Study site

The study site is on the southwest slopes of Mauna Kea Volcano on the Island of Hawaii (Figure 3.1) with trade wind inversion usually causing conditions on the site to be cool and dry. The elevation of the site ranges between 1285 m and 2662 m. The site is upland dry forest dominated by two tree species mamane and naio. The site has been heavily grazed by introduced ungulates, leading to the woodland now being degraded and scattered, with parts of the site having isolated trees. The study site straddles a fence that separates an unprotected area in the north of the study site from two protected areas in the south of the site. The two protected areas, a State Forest Reserve with open access in the east and a State Game Management Area with restricted access, requiring a hunting permit in the west are themselves separated by a fence (US Geological Survey (USGS) Gap Analysis Program (GAP), 2012). Several court cases most recently in 2009 have order that ungulates be removed from the site (Henkin, 2009). The site experienced a long-term drought between 2000–2011 (Banko et al., 2013). An initial survey of the site for palila began in the early 1980s, the survey technique was then changed, and survey effort increased in the 1999 survey year (Press et al. 2006).

Based on a survey conducted in the year 2000 around bird count stations, where the number of each tree species within a 40 × 40 m quadrat were counted at each station, the study site contained 93.6% mamane and naio trees out of the total number of trees (Brinck and Banko, 2014). The 2000 survey contained slightly more individual naio (54.04%) trees than individual mamane trees (46.0%). Mamane is the more important species for palila, although the palila will use both.
3.2 Data

The LiDAR data and hyperspectral data were both flown at the same time on the 9th of January 2009 around midday by The Carnegie Airborne Observatory (CAO) Beta system. This study used multipurpose survey station data. Three pieces of data were recorded in the palila count surrounding each station, tree survey data in 40 × 40 m quadrats in the year 2000 and the elevation of each station.

3.2.1 Hyperspectral data

The hyperspectral data recorded 36 bands between 389.8 and 1050.6 nm, with full width at half maximum (FWHM) for each band ranging from 18.8 nm at the lower end and 18.6 nm at the upper end of the spectrum (Table 3.1). The data are at 0.56 m resolution.
<table>
<thead>
<tr>
<th>Band</th>
<th>Wavelength (nm)</th>
<th>FWHM (nm)</th>
<th>Notable wavelengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>389.8</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>408.6</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>427.5</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>446.4</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>465.3</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>484.2</td>
<td>18.8</td>
<td>Blue</td>
</tr>
<tr>
<td>7</td>
<td>503.1</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>522.0</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>540.9</td>
<td>18.8</td>
<td>Green</td>
</tr>
<tr>
<td>10</td>
<td>559.8</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>578.7</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>597.6</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>616.5</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>635.4</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>654.3</td>
<td>18.8</td>
<td>Red</td>
</tr>
<tr>
<td>16</td>
<td>673.2</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>692.1</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>711.0</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>729.9</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>748.8</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>767.6</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>786.5</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>805.4</td>
<td>18.8</td>
<td>Infra-red for NDVI</td>
</tr>
<tr>
<td>24</td>
<td>824.3</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>843.2</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>862.1</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>880.9</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>899.8</td>
<td>18.8</td>
<td>Stable reference for WBI</td>
</tr>
<tr>
<td>29</td>
<td>918.7</td>
<td>18.8</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>937.5</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>956.4</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>975.2</td>
<td>18.6</td>
<td>Water absorption band WBI</td>
</tr>
<tr>
<td>33</td>
<td>994.1</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>1012.9</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>1031.7</td>
<td>18.6</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>1050.6</td>
<td>18.6</td>
<td></td>
</tr>
</tbody>
</table>
3.2.2 LiDAR data

The LiDAR system was operated in discrete capture mode capable of capturing up to four returns. The maximum number of returns collected was three. The data contained classes as defined by the American Society for Photogrammetry and Remote Sensing (American Society for Photogrammetry and Remote Sensing Standard, 2013; Esri, 2013), the classes in the LiDAR data and the number of points in each can be seen in Table 3.2. The data also contained the return number, the total number of returns and the intensity of the return ranging between 0 and 10750 (DN).

<table>
<thead>
<tr>
<th>Class</th>
<th>Point count</th>
<th>Percentage of total points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclassified (1)</td>
<td>5,796,775</td>
<td>3.7%</td>
</tr>
<tr>
<td>Ground (2)</td>
<td>6,502,631</td>
<td>4.1%</td>
</tr>
<tr>
<td>Low Vegetation (3)</td>
<td>91,229,156</td>
<td>57.9%</td>
</tr>
<tr>
<td>High Vegetation (4)</td>
<td>54,168,360</td>
<td>34.3%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>157,696,922</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

3.2.3 Stations

The stations extend all the way around Mauna Kea. The stations were used for multiple purposes. All stations contain a recorded elevation. The stations used for the analysis that contain palila data and surrounding zero values and which fall within the area containing both LiDAR and hyperspectral data (Figure 3.2) consists of 113 stations. The data contains counts of palila for 1998 through to 2009. This study will use the year 2009, which corresponds to the remotely sensed data that was collected. The mean elevation used for the station elevations ranged between 2122 m and 2631 m.
The remotely sensed area contained 41 stations with palila counted and 72 stations with no observed palila during the 2009 survey period. Within the boundary of the remotely sensed data, a maximum of five palila were observed at any one station. The total number of palila observed in 2009 increased from 2006 and 2007, but 2009 was still the third worst year within the period of the increased survey effort shown by these data. Figure 3.3 shows the higher numbers of palila observed in the late 1990s and early 2000s, with declines in the latter half of the 2000s.

Another use of the network of stations was for the locating of $40 \times 40$ m quadrats used to count how many trees of each tree species were present as well as characteristics of each tree and its surroundings in a 2000 survey. The tree survey did not use all of the stations used in the bird count. This survey used 128 of the stations in the network, 120 of which are within the extent of the LiDAR data. Between 3 and 150 individual trees were recorded in each quadrat. The precise location and orientation of each quadrat was not recorded in the data or in a subsequent report (Brinek and Banko, 2014).
Figure 3.3. Palila observed at each station for years of increased survey effort.
Table 3.3. Showing the variables to be considered in the analysis as well as the code used in R and the results (the codes will have Ma, Na and all at the end, which signify mamane, naio and combined), hypothesised sign of coefficient within the model and the relevance of the variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Code</th>
<th>Hypothesised sign of coefficient</th>
<th>Relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree count</td>
<td>Trees</td>
<td>Positive</td>
<td>Tree count is related to more biomass and therefore food, as well as diversity of fruiting.</td>
</tr>
<tr>
<td>Area of woodland intersecting buffered stations</td>
<td>SumAreaPatches</td>
<td>Positive</td>
<td>A larger area of woodland should contain more biomass and therefore food.</td>
</tr>
<tr>
<td>Mean area of contiguous canopy</td>
<td>MeanPatchArea</td>
<td>Positive</td>
<td>Mean area represents patch size.</td>
</tr>
<tr>
<td>Mean elevation of buffered station and</td>
<td>MeanElev</td>
<td>Positive</td>
<td>Elevation is related to pest species, with fewer found at higher elevation.</td>
</tr>
<tr>
<td>Total tree volume within buffered station area</td>
<td>SumVolume</td>
<td>Positive</td>
<td>A larger volume of trees should contain more biomass and therefore food.</td>
</tr>
<tr>
<td>Mean tree height within buffered station</td>
<td>MeanHeight</td>
<td>Positive</td>
<td>Tree height is thought to be related to tree age.</td>
</tr>
<tr>
<td>Mean ground intensity within buffered station.</td>
<td>MeanGrnInt</td>
<td>Negative</td>
<td>Intensity should be related to density of crown and the amount of browsing.</td>
</tr>
<tr>
<td>Mean Water Band Index (WBI) within buffered station</td>
<td>MeanWBI975</td>
<td>Positive</td>
<td>WBI should give water content of tree, as well as indication of locally available water.</td>
</tr>
</tbody>
</table>

3.3 LiDAR and hyperspectral data processing.

To allow for the testing of the hypothesised relationship of variables with palila, the data needed to be processed and variables derived from the remotely sensed LiDAR and hyperspectral data. This was achieved using ESRI ArcGIS version 10.2.2., ENVI 5.0, R and Microsoft Excel 2010.

The original raw LiDAR data consisted of each point having horizontal location (spatially), elevation, intensity of the return, the order of the return (first, second or third), and how many returns total were returned from the original pulse. The LiDAR points also contained a classification of what the point was likely to have hit. A digital terrain model (DTM) was needed to ascertain the mean elevation of buffered stations and was used in the creation of a canopy height mode (CHM).
The CHM gives the height of every raster pixel above the ground (the DTM). To obtain a count of the number of trees, the height of individual trees, and to identify mamane and naio individually, it was necessary to delineate individual tree canopies. These canopies were classified into individual tree species, using the intensity of the canopy returns and normalised difference vegetation index (NDVI) from the hyperspectral data, with each tree having implicit dimensions. Also of interest was the intensity of the ground returns from beneath the canopy.

The LiDAR data was converted from the LASer (LAS) File Format Exchange file format to an ESRI point feature class to allow for exploration within ArcScene and to ease subsequent manipulation. The LiDAR data has an average point spacing of approximately 0.21 m that increased towards the edges of the flown area to approximately 0.31 m. The LiDAR data already contained a classification as shown in Table 3.2. To allow for the combination of datasets derived from the LiDAR and the imagery all digital terrain models (DTM), canopy height models and interpolated intensity rasters, derived from the LiDAR point data, were snapped to the hyperspectral data and were created at the resolution of the hyperspectral data which was at a 0.56 m resolution.

3.3.1 Digital Terrain Model

The LiDAR point cloud was examined and the classes already within the data were deemed sufficiently good to preclude any reclassification. In the creation of the DTM, classes 2 and 3 (ground and low vegetation) were used together, due to the small number of ground returns alone as can be seen in Table 3.2. The point spacing of class 2 and 3 together was ~0.31 - ~0.37 m. With the point spacing being less than the intended final terrain model, a simple interpolation using inverse distance weighting was used, more complex Kriging was not necessary. The DTM was validated by comparing elevation recorded at the survey stations with elevation in the DTM.
**3.3.2 Canopy Height Model**

A canopy height model (CHM) is usually calculated as the difference between a DTM and a digital surface model (DSM). This was not the approach taken by this work, as the DSM had no further use in this study. The CHM was created by calculating the distance of LiDAR points to the DTM. This allowed for filtering of the points, using only those that came from the trees. Any point less than one meter above the DTM was assigned a new value of zero meters height. The filtered points were not interpolated. Interpolation would have brought the height of the canopy down, as it would have included points from within each canopy. Instead the points were directly converted to a raster, with each pixel being assigned the value of the highest LiDAR point within that pixel, further removing points that were below the top of the canopy. Any holes, pixels that had had no points over 1 m height within the canopies at 0.56 m resolution, were filled using the mean of the surrounding pixels.

**3.3.3 Tree isolation**

The next step in identifying individual tree canopies was the calculation of local maximum height of each tree. Due to multiple maxima within one canopy or smaller trees on the edges of larger canopies, this is not a simple procedure. The first option trialled was a variable distance search windows, based on the height of each pixel and derived allometric equations for canopy width at a particular tree height. This however led to the inclusion of many more points than were likely to be the number of trees, due to the search window being smaller when the height was less and the inclusion of many low local maximum around the edge of canopies. A second approach was employed using “focal flow”, which identifies cells that do not have any other cell flowing into it (therefore the peaks), which negates the need for a search radius. This process was conducted iteratively using the mean focal statistics with a two cell radius circular neighbourhood gradually smoothing the CHM. To ascertain the stopping point of the iterations three areas were manually
digitised from the hyperspectral data. The manually digitised trees were likely incorrect themselves as distinguishing trees in the hyperspectral data was difficult or impossible in many cases. The iteration was chosen that visually gave a point in each of the digitised trees, airing on the side of caution, preferring too many points than too few, as this is less likely to clump different tree species together accidentally. Two smoothing iterations were used. Figure 3.4 shows a mock-up of the iterative smoothing. The smoothed CHM was only used to identify a point within each canopy. Having identified a single point within a canopy the original CHM was used to create each delineated canopy using a form of watershed delineation. The smoothed CHM would have smoothed the boundary between each canopy and would give incorrect height and volume measurements.

![Figure 3.4](image)

**Figure 3.4.** Shows an example of the smoothing process. a) shows hypothetical CHM with multiple local maximum. b) shows intermediate smoothed canopy, still showing two local maximum. c) shows final smoothed CHM, with the focal flow now giving multiple points adjacent to each other.

The unsmoothed CHM was inverted and the values made positive, to form basins. The inverted CHM was then exponentially increased to increase the difference between individual canopies. A cost distance analysis was performed using the identified maxima. Watersheds were then calculated on the cost distance layer again from the local maxima. These watersheds represent individual canopies.

As validation, the individual trees were compared to the 2000 tree survey. The precise locations of the 40 × 40 m quadrats were unknown; therefore, the stations were buffered using a radius of 22.6 m to give an area equal to the quadrats (1600 m²).
3.3.4 Tree species mapping

Using the watershed delineated individual canopies it was possible to classify the tree species of individual canopies. The edges of individual canopies are often shadowed, or are mixed with ground pixels. To avoid noise from mixing with ground pixels or from a low signal to noise ratio in shadows within edge pixels, one cell width (0.56m) from the edge of each canopy was removed, if the tree was large enough to do so. Canopies are also often intimately mixed with neighbouring trees with the area between them potentially reflecting from two different tree species. To try to remove mixing from neighbouring trees, convex hulls of each canopy were created and where these convex hull polygons overlapped these overlapping areas were also removed within GIS by intersecting the polygons, growing the overlapping region by one pixel, and then erasing this area from the tree canopy polygons.

Given the dominance of mamane and naio the classification only attempted to classify trees into either mamane or naio, accepting that some other species, although very few, would be classified wrongly as either mamane or naio. An area of known conifer plantation was removed manually. The numbers of dead trees are also thought to be few. Looking at two areas of trees in the hyperspectral data, which from the 2000 tree survey contained 100% mamane and 94% naio, the spectral signature of the trees in these areas are virtually identical. Figure 3.5 shows the spectrum for the two areas and for comparison, an area of plantation, which is very different.
Figure 3.5. Spectral signature of naio dominated areas in purple (the bottom of the two close lines), mamane dominated area in green (the top of the two close lines) and an area of conifer plantation in red (the line most different from the other two).

Normalized difference vegetation index (NDVI) can be calculated using the red band (Red = 635.4 nm) and a near infra-red band (NIR = 805.4 nm):

\[
NDVI = \frac{(NIR - Red)}{(NIR + Red)}
\]

The NDVI had different means for the two areas (0.16 and 0.17), with mamane dominated area having the higher mean NDVI. NDVI for all trees however was normally distributed and mamane and naio were indistinguishable.

The mean LiDAR canopy return intensity had a bimodal histogram. Separation using a single value cut off between the two modes would however lead to misclassification in the area where the two distributions overlap. K-mean clustering in R was therefore used, using the mean LiDAR canopy return intensity and NDVI together. Using only two clusters, examined initially, led to a classification similar to using a cut-off value. Increasing the number of clusters to many more than needed (twenty were used), and then manually assigning each to one of the two distributions of mean LiDAR canopy return intensity, led to the two distinct histograms shown in Figure 3.6. Some
misclassification is still present, evident in the sharp vertical edges of the two histograms, but the proportion of mamane and naio is sensible, 48% mamane, 52% naio.

![Histogram of intensity of LiDAR top of canopy returns](image)

**Figure 3.6.** Showing a histogram of intensity of LiDAR top of canopy returns, showing clearly two different groupings of intensity.

As validation, the percentage of mamane trees from the all trees were compared to the 2000 tree survey using the buffered stations with an area of 1600 m$^2$ described at the end of the last section.

### 3.3.5 Individual tree size

Tree height was calculated as the highest pixel from the CHM within each individually differentiated tree canopy. For simplicity volume under the canopy of each tree was used instead of true canopy volume, we did not seek to find the base height as it is likely not worth the effort that would be involved and a simple volume is sufficient for comparison between trees within different buffers of each station.
3.3.6 Canopy density

The current study made a relative measure of canopy density but did not calibrate the exact density using ground truthing. This relative measure should be sufficient as a co-variable. Mean ground return intensity below each canopy was calculated by first creating an interpolated raster of the intensity of the ground return points (class 2 and 3) and then calculating the mean value under each canopy. Mean ground return intensity was also graphed against tree height to explore whether the height of individual trees influenced the mean ground intensity below the trees.

Validation of the canopy density is troublesome as field data was not collected in 2009. Field measurements of canopy density or browsing therefore do not exist. However, it was possible to compare inside and outside of the protected area fence. It is not possible with the data available to control for different influences inside and outside the fence, but browsing is uncontrolled outside the fence. Outside the fence will, therefore, be treated as higher browsing and inside as lower browsing, with no area having no browsing.

3.3.7 Patch size

The current analysis treated each individual station as a separate set of observations, therefore measuring the width of a wider woodland encompassing multiple stations, as had been done previously, would not be appropriate at this scale. This correlation of palila with width of woodland suggests that a measure of patch size could be important to palila. To try and capture patch size, the mean area of the each distinct contiguous patch of tree canopy within the buffer around the stations was included.

3.3.8 Elevation

The mean of the elevation from the DTM was taken within the different buffers of each station. The buffered survey stations had a mean elevation range of just over 500 m.
3.3.9 Water content

Using the individual mamane and naio trees, it was possible to compare the spectral signatures of each tree and calculate a relative water stress or water content. The longest wavelength in the hyperspectral data used in this study is 1050 nm. The water band index (WBI) (Peñuelas et al., 1993) uses two bands both within the sites hyperspectral data, 900 nm and 970 nm:

\[ WBI = \frac{R_{900}}{R_{970}} \]  

R is the reflectance at wavelengths 900 and 970 nm. WBI will give a relative measure of the water content of each canopy. As the water content increases the 970 nm band increases relative to the 900 nm band, therefore higher values of the ratio of 900 nm divided by 970 nm have more water content (EXELIS visual information solutions, 2014; Peñuelas et al., 1993).

To validate the WBI, interpolated monthly rainfall will be used for comparison (Frazier et al., in review.) The rainfall data is at 8.1 arcseconds resolution (∼236 × 248 m); to allow for comparison the WBI for all mamane and naio was averaged for pixels of the same size to allow for comparison of the general trend.

3.4 Scale of analysis around bird count stations.

Each station containing bird count data is a single point. To allow for comparison of the habitat variables an area around the stations must be used to amalgamate the variables. The literature suggested that a grain size similar to that of the home range of the bird should give the best result. Assuming this is a circular home range and using home range from Farmer et al. (2014), the radius is then approximately 180 m ± ~3 m. Therefore, a 180 m buffer of the stations should best explain the 2009 palila count. This study will compare this 180 m buffer around each station to larger (360 and 900 m) and smaller (90 m) buffer sizes.

3.5 Linear regression and analysis

All model creation and analysis was conducted using the statistical package R (R Core Team, 2013). Multiple negative binominal linear regression was used due to Poisson regression, which can
be used for regression analysis when the dependant variable is count data, being overdispersed (the variance of the data was greater than the mean). Poisson regression was therefore not appropriate, negative binominal regression is similar to Poisson regression but can deal with overdispersion.

Equations were set up using the MASS package (Venables and Ripley, 2002) comparing the 2009 palila count to the variables in Table 3.3 for: i) mamane trees, ii) naio trees, iii) both mamane and naio trees separately iv) combined for all trees (mamane and naio together), and v) all mamane and naio variables separately and combined.

Bayesian information criterion (BIC) stepwise selection (AIC stepwise with k = ln(n), where n is the number of records; n = 113, therefore k ~ 4.7) in both directions (Venables and Ripley, 2002) will be used to whittle the variables down to the variables with the most explanatory power in explaining palila numbers. BIC was used instead if AIC as BIC applies a higher penalty, decided based on the number of records, for more variables. BIC lead to more robust significant models than AIC. All analysis were run with buffers of 90, 180, 360 and 900 m radius and the models collectively compared to ascertain the validity of using the home range (180 m buffer). Equation 1 and 2 will be compared to see if mamane trees alone explain the palila count better than naio. Equation 3 and 4 will be compared to see if combined measures are more appropriate than separate variables.

As a means of comparing the model, Bayesian information criterion (BIC) (Schwarz, 1978), McFadden’s pseudo $R^2$ (1 - log likelihood(model)/log likelihood(null)) (Fultz, 2012) and Leave One Out Cross Validation (LOOCV) estimated prediction error (using the BOOT package) (Canty and Ripley, 2014; Davison and Hinkley, 1997) was calculated for each model. The most sensible model will be selected, using BIC, Pseudo $R^2$ value and leave one out cross validation (LOOCV). If the difference between the BICs of different models is more than 2, the difference is significant, greater than 6 is strongly significant, and over 10 very strongly significant (Kass and Raftery, 2015). Pseudo
$R^2$ values closer to one suggest a more likely model (Fultz, 2012). The lower the LOOCV estimated prediction error, the better the model would be at predicting with new data. The variables in this model that best explain the 2009 palila count will then be examined to see if they agree with the hypothesised sign of the coefficient. Each will be examined separately against the count data to check that the sign is not a result of interactions between the variables.
Chapter 4.
RESULTS

4.1 Validation

4.1.1 Digital terrain model

Plotting elevation recorded at the survey stations against that estimated using the LiDAR derived DTM the two are highly correlated (Figure 4.1) with an $R^2$ of 0.9994 and a slope of almost one. The height is, however, approximately 25 m higher at all location than that recorded in the survey stations.

![Figure 4.1. Comparison of elevation recorded at survey stations and from the DTM](image)

\[ y = 0.9932x + 41.068 \]
\[ R^2 = 0.9994 \]
4.1.2 Tree isolation

Plotting the 2000 data against the count from the canopy height model (CHM) delineated tree canopies within the 1600 m$^2$ circular buffers, shown in Figure 4.2, shows that the delineation appears to underestimate trees when the 2000 survey counted high numbers of trees.

![Figure 4.2. Count of trees in the year 2000 survey, against count from the delineation in this study.](image)
Looking at the station with the highest count from the 2000 survey shown in Figure 4.3 the high counts of the 2000 survey are difficult to see in the figure. For a visual comparison of every station, see Appendix 3.

4.1.3 Tree species

Plotting percentage of mamane from all trees in the 2000 tree survey against the percentage of mamane from all trees in the 1600 m² buffers from the CHM delineated canopies, shows
considerable over and under estimation of the percentage (Figure 4.4). The general pattern is correct in most cases (93% are within 50 percentage points), with mamane or naio dominated areas still dominated by mamane or naio.

Figure 4.4. Percentage of mamane from total number of trees in the 2000 survey and the estimation from the current survey. Black line represents 1:1 line also included count of over, equal and under estimation and RMSE.

**4.1.4 LiDAR ground return intensity**

Figure 4.5 shows mean LiDAR ground return intensity (MLGRI) below mamane and naio canopies at binned 0.25 m height intervals, outside of the protected areas fence and for the two protected areas that are separated by a fence. In Figure 4.5 MLGRI under mamane is clearly higher outside the protected areas fence, more of the LiDAR pulse is being transmitted through the tree
canopy and is being reflected to the LiDAR sensor. There is much less difference between the two fenced protected areas. The intensity becomes similar again when the mamane trees are taller. The relationship appears to be similar for naio trees, although only as clearly for trees between approximately 3 and 10 m.
Figure 4.5. LiDAR ground return intensity under mamane and naio tree canopies, outside of protected areas and fence, and in the two protected areas, showing intensity across a range of heights.
4.1.5 Water content

Figure 4.6 and Figure 4.7 plots interpolated inches of rainfall (Frazier et al., in review.) against the mean water band index (WBI), scaled up to the same resolution as the rainfall data for mamane and naio trees separately.

Looking at Figure 4.6 WBI appears to track well with rainfall for mamane trees below approximately 4.2 inches of rain, above this, the WBI varies. It could potentially be levelling off at around a WBI of 1.06, although another horizontal line could be drawn through a WBI of 0.99.

The WBI of the Naio trees tracks reasonable well below a rainfall of 4 inches when compared to the December 2008 interpolated rainfall (top Figure 4.7). The points above 4.3 inches and a WBI of 1.05 (top right) are in the very far north of the site, with extremely few naio trees. The points above 4.15 inches and below a WBI of 1.025 (bottom right) are, with the exception of two points which are again in the far north of the site, in the southwest edge of the site. The values around the hand drawn black line are from areas dominated by mamane north of a line (shown in blue diamonds in both graphs in Figure 4.7). Comparing the WBI to the interpolated January 2009 rainfall (bottom Figure 4.7) most of the points from the December
2008 comparison now appear to fit the expected positive relationship. The area that did not fit in December 2008, with high rainfall, had the lowest rainfall in January 2009.

Figure 4.7. Mean WBI for naio trees compared to interpolated inches of rainfall in December 2008 and January 2009 from Frazier et al., (in review.) The data has been split into two broad groups, which either fit (blue diamonds) or do not (red circles) fit a relationship of rainfall in the 2008 data. The black lines are hand drawn showing a line of fit.
4.2 Analysis

Table 4.1 shows the resulting models of the stepwise selection. The first seven columns relate to the whole model, giving the name of the model, the buffer distance of each grain size, Bayesian information criterion (BIC) value, delta BIC, McFadden’s pseudo $R^2$, estimated prediction error from leave one out cross validation (LOOCV) and the intercept. Within Table 4.1 the last three columns give the coefficients of the model and the estimated standard error and p-value for each coefficient. There are different grain sized variables of each model, the for each model and the difference in BIC from the model with the lowest BIC.

4.2.1 Scale of analysis

For the models created using Bayesian information criterion (BIC) stepwise selection, shown in Table 4.1, BIC is similar for all models. The lowest BIC is for the 180 m radius grain size separate mamane and naio, and all variables models, both having selected the same variables. The BIC is within two for the best models within each buffer distance, the pseudo $R^2$ values are similar, and the estimated prediction errors from leave one out cross validation (LOOCV) are similar. The variables selected at these different scales vary.

4.2.2 Mamane vs. naio trees

Looking at the models in Table 4.1, there is no significant difference between the BIC for mamane or naio at most scales. The exception being the 900 m buffer distance which threw out all variables in the naio model, giving the null model. The pseudo $R^2$ and estimated prediction error from LOOCV are also similar.

4.2.3 Combined vs. separate tree species

In the 90 m radius buffer distance, the combined and separate models have a BIC less than two different (Table 4.1). The separate model does have the lower BIC and has a higher pseudo $R^2$ and lower estimated prediction error from LOOCV. At all other buffer distances the combined model was the null model, not including any variables.
Table 4.1. BIC stepwise \( k = 4.7 \) selected negative binominal linear regression models for different scales, and with different starting variables.

<table>
<thead>
<tr>
<th>Model</th>
<th>Buffer radius</th>
<th>BIC</th>
<th>ΔBIC</th>
<th>Pseudo R²</th>
<th>Est. prediction err</th>
<th>Intercept</th>
<th>Regression coefficient</th>
<th>Est. Std. err</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mamane</td>
<td>90</td>
<td>258.95</td>
<td>1.65</td>
<td>0.90</td>
<td>1.17</td>
<td>23.69</td>
<td>-0.0909 (MeanGrnIntMa)</td>
<td>0.0343</td>
<td>0.00801</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-18.3 (MeanWBI975Ma)</td>
<td>7.54</td>
<td>0.01515</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+0.696 (MeanHeightNa)</td>
<td>0.232</td>
<td>0.00273</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0635 (MeanGrnIntNa)</td>
<td>0.0284</td>
<td>0.02533</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+0.00398 (MeanElev)</td>
<td>0.00161</td>
<td>0.01360</td>
</tr>
<tr>
<td>Naio</td>
<td>90</td>
<td>259.57</td>
<td>2.27</td>
<td>0.86</td>
<td>1.17</td>
<td>-11.48</td>
<td>-0.0732 (MeanGrnIntAll)</td>
<td>0.0265</td>
<td>0.00573</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-28.7 (MeanWBI975All)</td>
<td>11.2</td>
<td>0.01038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+0.00409 (MeanElev)</td>
<td>0.00169</td>
<td>0.01565</td>
</tr>
<tr>
<td>Combine</td>
<td>90</td>
<td>260.77</td>
<td>3.47</td>
<td>0.87</td>
<td>1.21</td>
<td>23.07</td>
<td>-0.0909 (MeanGrnIntMa)</td>
<td>0.0343</td>
<td>0.00801</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-18.3 (MeanWBI975Ma)</td>
<td>7.54</td>
<td>0.01515</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0635 (MeanGrnIntNa)</td>
<td>0.0284</td>
<td>0.02533</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+0.00398 (MeanElev)</td>
<td>0.00161</td>
<td>0.01360</td>
</tr>
<tr>
<td>Separate</td>
<td>90</td>
<td>258.95</td>
<td>1.65</td>
<td>0.90</td>
<td>1.17</td>
<td>23.69</td>
<td>-0.0909 (MeanGrnIntMa)</td>
<td>0.0343</td>
<td>0.00801</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-18.3 (MeanWBI975Ma)</td>
<td>7.54</td>
<td>0.01515</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0635 (MeanGrnIntNa)</td>
<td>0.0284</td>
<td>0.02533</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+0.00398 (MeanElev)</td>
<td>0.00161</td>
<td>0.01360</td>
</tr>
<tr>
<td>All</td>
<td>90</td>
<td>259.21</td>
<td>1.91</td>
<td>0.86</td>
<td>1.13</td>
<td>-0.95</td>
<td>-0.000057 (MeanPatchAreaNa)</td>
<td>0.000263</td>
<td>0.02990</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+0.522 (MeanHeightNa)</td>
<td>0.227</td>
<td>0.01240</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0443 (MeanGrnIntMa)</td>
<td>0.0175</td>
<td>0.01120</td>
</tr>
<tr>
<td>Mamane</td>
<td>180</td>
<td>259.63</td>
<td>2.33</td>
<td>0.95</td>
<td>1.17</td>
<td>2.92</td>
<td>-0.0579 (MeanGrnIntMa)</td>
<td>0.0257</td>
<td>0.02450</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+0.0635 (MeanGrnIntNa)</td>
<td>0.0261</td>
<td>0.01500</td>
</tr>
<tr>
<td>Naio</td>
<td>180</td>
<td>258.55</td>
<td>1.25</td>
<td>0.94</td>
<td>1.14</td>
<td>2.18</td>
<td>-0.0635 (MeanGrnIntMa)</td>
<td>0.0261</td>
<td>0.01500</td>
</tr>
<tr>
<td>Separate</td>
<td>180</td>
<td>257.30</td>
<td>0</td>
<td>0.88</td>
<td>1.18</td>
<td>3.25</td>
<td>+0.0000905 (SumAreaPatchesMa)</td>
<td>0.0000376</td>
<td>0.1603</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.114 (MeanGrnIntNa)</td>
<td>0.038</td>
<td>0.00265</td>
</tr>
<tr>
<td>All</td>
<td>180</td>
<td>257.30</td>
<td>0</td>
<td>0.88</td>
<td>1.18</td>
<td>3.25</td>
<td>+0.0000905 (SumAreaPatchesMa)</td>
<td>0.0000376</td>
<td>0.1603</td>
</tr>
<tr>
<td>Mamane</td>
<td>360</td>
<td>258.97</td>
<td>1.67</td>
<td>0.94</td>
<td>1.17</td>
<td>3.46</td>
<td>-0.0672 (MeanGrnIntMa)</td>
<td>0.0283</td>
<td>0.01740</td>
</tr>
<tr>
<td>Naio</td>
<td>360</td>
<td>258.78</td>
<td>1.48</td>
<td>0.94</td>
<td>1.14</td>
<td>2.52</td>
<td>-0.0729 (MeanGrnIntNa)</td>
<td>0.0302</td>
<td>0.01600</td>
</tr>
<tr>
<td>Separate</td>
<td>360</td>
<td>257.38</td>
<td>0.08</td>
<td>0.84</td>
<td>1.15</td>
<td>65.98</td>
<td>+0.0000374 (SumAreaPatchesMa)</td>
<td>0.0000144</td>
<td>0.0000934</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.394 (MeanGrnIntNa)</td>
<td>0.13</td>
<td>0.00241</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.114 (MeanGrnIntMa)</td>
<td>0.038</td>
<td>0.00265</td>
</tr>
<tr>
<td>All</td>
<td>360</td>
<td>261.53</td>
<td>4.23</td>
<td>0.72</td>
<td>3.76</td>
<td>108.60</td>
<td>+0.000273 (SumAreaPatchesMa)</td>
<td>0.0000929</td>
<td>0.00326</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+2.09 (MeanHeightMa)</td>
<td>0.697</td>
<td>0.00277</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0000794 (SumVolumeMa)</td>
<td>0.0000267</td>
<td>0.00296</td>
</tr>
<tr>
<td>Mamane</td>
<td>900</td>
<td>258.08</td>
<td>0.78</td>
<td>0.89</td>
<td>1.17</td>
<td>1.37</td>
<td>+0.120 (MeanPatchAreaMa)</td>
<td>0.0444</td>
<td>0.00668</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0957 (MeanGrnIntMa)</td>
<td>0.0425</td>
<td>0.02434</td>
</tr>
</tbody>
</table>
4.2.4 Explanatory variables

The hypothesis is that: the model that best explains the 2009 palila count, will include the following variables, the order below represents the hypothesised importance:

a. WBI
b. Total volume
c. Mean LiDAR ground return intensity
d. Elevation
e. Mean contiguous canopy

Area covered by trees and tree height will be unimportant predictors.

Also hypothesised: that with the exception of Mean LiDAR ground return intensity, all the coefficients are expected to be positive.

Table 4.2. Coefficients in single variable models at different radius grain size.

<table>
<thead>
<tr>
<th></th>
<th>90 m</th>
<th>180 m</th>
<th>360 m</th>
<th>900 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>MeanGrnIntMa</td>
<td>-0.0485</td>
<td>-0.0579</td>
<td>-0.0672</td>
<td>-0.06325</td>
</tr>
<tr>
<td>MeanGrnIntNa</td>
<td>-0.00511</td>
<td>-0.0635</td>
<td>-0.0729</td>
<td>-0.0883</td>
</tr>
<tr>
<td>MeanGrnIntAll</td>
<td>-0.0247</td>
<td>-0.0273</td>
<td>-0.0295</td>
<td>-0.0337</td>
</tr>
<tr>
<td>MeanWBI975Ma</td>
<td>-7.55</td>
<td>0.00497</td>
<td>3.02</td>
<td>2.75</td>
</tr>
<tr>
<td>MeanWBI975Na</td>
<td>3.86</td>
<td>9.44</td>
<td>9.12</td>
<td>19.1</td>
</tr>
<tr>
<td>SumAreaPatchesMa</td>
<td>0.000125</td>
<td>0.0000276</td>
<td>0.00000581</td>
<td>0.00000114</td>
</tr>
<tr>
<td>SumAreaPatchesNa</td>
<td>-0.00000615</td>
<td>-0.000000634</td>
<td>0.000000270</td>
<td>0.000000365</td>
</tr>
<tr>
<td>SumAreaPatchesAll</td>
<td>-0.00000111</td>
<td>-0.00000114</td>
<td>0.00000116</td>
<td>0.00000497</td>
</tr>
<tr>
<td>MeanPatchAreaMa</td>
<td>0.0307</td>
<td>0.0249</td>
<td>0.0305</td>
<td>0.0922</td>
</tr>
<tr>
<td>MeanPatchAreaNa</td>
<td>-0.000222</td>
<td>-0.000168</td>
<td>0.000257</td>
<td>0.00213</td>
</tr>
<tr>
<td>MeanHeightMa</td>
<td>0.00705</td>
<td>-0.320</td>
<td>-0.347</td>
<td>-0.130</td>
</tr>
<tr>
<td>MeanHeightNa</td>
<td>0.427</td>
<td>0.163</td>
<td>-0.109</td>
<td>-0.142</td>
</tr>
<tr>
<td>MeanElev</td>
<td>0.00232</td>
<td>0.00238</td>
<td>0.00277</td>
<td>0.00284</td>
</tr>
<tr>
<td>SumVolumeMa</td>
<td>-7.55</td>
<td>0.00497</td>
<td>3.02</td>
<td>2.75</td>
</tr>
</tbody>
</table>

The p-value was significant for all variables selected above 97% confidence, with most above 99%. None of the significant models contains multicollinearity. In every model, MLGRI below the tree canopy was included in the model. In the models with the lowest BIC, estimated prediction error from LOOCV and highest pseudo $R^2$ within each grain size, the intensity was that below either, the mamane or naio trees, not a combined measure for both tree species. The combined measure was selected in other models that are not significantly models. The coefficient
in every case was negative. When looking at the intensity alone as a predictor of the 2009 palila count (Table 4.2), the coefficient was consistently negative across all scales and tree species.

The WBI was included in two significant models. At the 90 m radius grain size, for mamane and for 360 m radius grain size for naio, although this model included one more coefficient than the 90 m model. In all models, the coefficient was negative. When looking at the WBI alone as a predictor of the 2009 palila count (Table 4.2), the coefficient was positive in most cases and only negative for the mamane trees at a 90 m radius grain size. The coefficients in these models were not significant, and in all cases where the coefficient was positive, the standard error of the coefficient was larger than the value was positive. This means that the coefficient could be negative. In the case of the 90 m mamane WBI model, the standard error of the coefficient was a smaller values than the coefficient was negative, therefore the coefficient is likely negative.

The area of mamane was present in two unique significant models, at a 180 and 360 m radius grain size. Area of naio or all trees were only found in one model with a standard error of 3.76 at a 360 m radius grain size, the model was insignificant. In all cases, area of mamane or all trees had a positive coefficient. In the insignificant model with area of naio, the coefficient was negative. The mamane area consistently had a positive coefficient when looking at the area of mamane trees alone as a predictor of the 2009 palila count (Table 4.2), the coefficient was not consistent for the area of naio or all trees combined, negative at 90 and 180 m and positive at 360 and 900 m.

The mean patch size of mamane and naio were found in one model each, at 900 and 90 m respectively. The mamane coefficient was positive and the naio coefficient negative. When looking at mean patch size alone as a predictor of the 2009 palila count (Table 4.2), the coefficient had the same pattern as the total areas, the mamane coefficient was positive; the naio coefficient was negative at 90 and 180 m and positive at 360 and 900 m. The mean patch sizes
and sum of area are highly correlated, 0.77 for mamane and 0.93 for naio. There are no models that contain both mean patch size and sum of area.

Mean naio height only appeared in one borderline significant model and one less significant model. Mean mamane height only appeared in one less significant model. In all cases, the coefficient was positive. Mean height was inconsistent when look at height alone as a predictor of the 2009 palila count (Table 4.2).

Elevation was present in two less significant models at 90 m radius grain size, each with a positive coefficient. Elevation was consistently positive when looking at elevation alone as a predictor of the 2009 palila count (Table 4.2).

Mamane volume only appeared in insignificant models, with a negative coefficient. Mamane volume was only negative at 90 m when looking and Mamane volume alone as a predictor of the 2009 palila count (Table 4.2), at all other grain sizes the coefficient was positive.

Tree count was not found in any model.
Chapter 5.

DISCUSSION

The purpose of this study was to assess which palila habitat variables, identified in previous studies, explains palila abundance in the 2009 palila count the best. This was achieved by using remotely sensed LiDAR and hyperspectral data as a tool to quantify habitat variables. These variables were validated and then used to create initial linear models at different scales, using different tree species and with different ways of representing combined tree species. Variables were removed using Bayesian information criterion (BIC) stepwise selection. This discussion will discuss the validation of the variables as well as potential improvements and refinements, go on to look at the regression analysis and any uncertainty in the analysis, the implications of the variables. Then the variables will be discussed in the context of the literature, and finally potential future work and improvements will be discussed.

The digital terrain model (DTM) was validated by comparing the recorded elevation at each station with that estimated in the DTM. The DTM predicted elevation almost perfectly; the elevation of the DTM was however approximately 25 m higher than that recorded in the survey station data. This is systematic error is from a difference in vertical datum, one giving orthometric height and the other geoid height.

As validation of the tree delineation, the numbers of trees delineated from the canopy height model (CHM) within a buffer around each station were compared to the number of trees in the quadrats in the 2000 tree survey, both quadrat and buffer had an area of 1600 m². Looking at the count data it appears that when the 2000 survey count was high the delineation underestimated the number of trees. Doing a visual comparison of each station, the number of canopies appears approximately correct. Presumably, the 2000 survey counted the number of stems at breast height. In these very high-count areas, each canopy likely contained multiple stems that contribute to
individual canopies. In places, it is hard to differentiate individual canopies manually. The continuous areas of canopy in some areas have a relatively flat canopy, which look similar in the hyperspectral data.

Looking at the proportion of mamane from all trees in the current work and in the 2000 survey, in most cases the current work both under and over estimates the percentage of mamane trees. The general trend is correct, with areas dominated by mamane still dominated by mamane. The discrepancy could be due to the difference in looking at individual stems in 2000 and at the overall canopy in the current study. If the 2000 survey had contained 20 mamane stems under one canopy and one naio, the percentage of mamane would be 95% mamane, in the current work that would be 50% mamane. The spectral signature in the hyperspectral data was very similar between mamane and naio. When looking at a known area of plantation, the signature was very different, with lower reflection of visible light and much higher near infra-red. Mamane and naio were very similar across all wavelengths. The 975 nm water absorption band was avoided as it would vary with the water content of the trees and was needed for the water band index (WBI) calculations. Given the similarity of the spectral signatures of the two tree species of interest, LiDAR intensity was explored as a means of separation. The mean canopy LiDAR intensity data had a bimodal distribution and allowed for separation of the two tree species. The two distributions had overlapping ranges, to allow for further separation the NDVI, which varied amongst trees, but was ineffective alone, was used to try and further separate the two tree species. In future studies full waveform may have allowed more information to be derived from the LiDAR data that could separate tree species. The discrete return LiDAR although capable of capturing up to four returns in most cases had only collected two, full waveform may have allowed for a more complete profile as well as more information on the LiDAR intensity.
Mean LiDAR ground return intensity (MLGRI) was created in an attempt to create a proxy for browsing. The MLGRI difference inside and outside the fence cannot categorically be stated as being a difference in browsing on the trees, and could include an element of the field layer (grass and other low-lying vegetation). The field layer, however, will also differ with browsing. The difference between mamane and naio does lend some support to the measurement of the trees themselves. Mamane and naio have different leaf structure, with naio leaves being larger. It is also thought that mamane are preferentially browsed (van Riper III, 1980). The MLGRI appears to vary with height, although less substantially inside the fenced protected areas. Ungulates are still present within the fenced area, but effort has been made to reduce their numbers, outside the fence, they are largely uncontrolled. The MLGRI does perform as expected, the MLGRI being higher where browsing is likely higher, outside the fence. For naio the MLGRI is not as clear outside the fence, but there are only few scattered naio outside the fence, therefore this is a mean of fewer trees, or even simply the value of an individual tree, and is likely a random values from the distribution around what the mean would be with more trees. Future research should look more into this measure, or a refined version, determining what exactly is influencing the difference in MLGRI, it should also be investigated whether this is applicable elsewhere in different types of forested habitats and not only similar woodlands with two species and limited understory.

The water band index (WBI) does appear to have a relationship with interpolated rainfall from Frazier et al., (in review.). This is particularly clear when looking at the mamane trees compared to rainfall in December 2008. Both the mamane and the two naio graphs of rainfall against WBI do have points that do not fit the relationship. For mamane, most of these points are from the extreme ends of the study site, areas that have few mamane trees. Similar to the MLGRI, these values are likely due to the value being more random from within a distribution around what the mean would be with more trees. It is also possible that the relationship is not linear and the WBI
does not increase beyond a value representing full water saturation. The explanations of the outliers from the mamane graph also hold true for some of the naio outliers when compared to rainfall in December 2008 and January 2009, but not all. Where there does appear to be a relationship between rainfall in December 2008 and naio WBI, the naio are all north of a line. Comparing the naio WBI to January 2009 (the hyperspectral data was collected on 9th of January 2009), the cluster of points that did not fit the December 2008 data now appear to fit the 2009 data. The rainfall across the site is less in January 2009 than December 2008, and areas amongst the wettest in 2008 were the driest areas in 2009. This difference when looking at December 2008 and January 2009 could be due to naio being affected by lack of rainfall quickly. This could make WBI less useful in predicting palila as the hyperspectral data would need to be collected very close to the bird count data in areas where the rainfall is very different from month to month. Palila counts usually take place in January (Camp et al., 2014), so in this case it is not too much of an issue. Having shown the relationship of the scaled up WBI to rainfall, it is possible that the WBI could somehow be used to improve spatial accuracy of rainfall interpolation, although the variability in tree species, slope, and infiltration would need to be accounted for. This suggests an avenue for future research. It may be more appropriate when trying to capture relatively rapid fluctuations in vegetation water content, or when hoping to add detail to interpolated rainfall to use satellite data with a relatively frequent return time, this would likely require the use of a different water absorption band, for example Landsat TM, does not collect in the 975 nm wavelength.

The other variables such as height, volume, and area were not validated, but were assumed approximately correct. Each was derived more directly from the LiDAR data and even if heights or areas were off a little, due to error from the LiDAR, on average a shorter site would still be shorter.

Having derived the variables of interest from the LiDAR and hyperspectral data, negative binominal linear regression equations were set up. The different equations were used to compare
scale of analysis, looking at 90, 180 (palila home range (Farmer et al. 2014)), 360 and 900 m. They were used to see if mamane variables better predicted palila in 2009 than naio variables and if having mamane and naio separately in a model, predicted better than combined. Finally the equations were used to look at all tree variables, and to identify which variables best explain the palila count data. Bayesian information criterion (BIC) stepwise selection (BIC applies a penalty derived from the number of records) was used to reduce the number of variables to those that most strongly explained the palila count. The BIC stepwise aggressively removed variables, reducing a few of the models to the null model, including no variables. This resulted in confidence in the selected surviving variables as being truly relevant.

Grain size was of interest in this study, what scale to best aggregate variables around the stations that were used for palila count. Meyer and Thuiller (2006) suggests that the best grain size to perform aggregation of variables when looking at habitat species relationships is that of the home range of the species. The home range grain size of palila from Farmer et al., (2014) has a radius of 180 m. This study also examined 90, 360, and 900 m radius grain size as a means of testing the validity of use of and validity of 180 m radius palila home range. The model with the lowest BIC was at a 180 m grain size. Models at the other grain sizes had BICs that were not significantly different and some had higher McFadden’s pseudo R² and lower estimated standard error. This suggests either that, the scale of analysis is unimportant, the true scale has not been identified or that different variables act at different scales, a multi-scale model would therefore be appropriate, something noted by Meyer and Thuiller (2006).

Palila are more closely associated with mamane than naio (Banko et al., 2002a), therefore it was hypothesised that mamane variables would best explain the 2009 palila count. Comparing models derived from only mamane variables and from only naio variables, there is little difference between mamane and naio models. The models with the lowest BIC and highest pseudo R² are each
mamane models, but the models that have the joint lowest estimated standard error are naio only models. The overall best model also only contains one variable, mean LiDAR ground return intensity (MLGRI) under naio canopy, this could be an extreme browsing measure, as areas that have browsing of naio, could have mamane that are heavily browsed (as mamane are thought to be preferred by ungulates) and maybe difficult to detect remotely. Overall models containing only mamane variables or naio variables perform the same.

Given that mamane and naio models are similar, it is possible that mamane and naio should not be separated and instead a combined all tree variable used. Looking at models containing separate mamane and naio, and combined all tree variables, separate models are clearly better. There is not a great deal of difference between the BICs when both models survive the BIC stepwise selection (only the case at 90 m radius grain size). The reason that the separate models are preferable is that for 180, 360 and 900 m radius grain size the combined model was the null model that only included the intercept. In the case of 900 m radius grain size the separate model was also the null model. An explanation of this could have been that the separate models could include more variables; however, comparing the models at 90 m radius grain size, the separate model also had fewer variables than the combined model.

Models containing both the separate mamane and naio, and combined trees variables were also included in a model labelled “All” to allow for all combinations of variables. No significant model ended up including more than one mamane, naio, or combined version of a variable. Given the similarity in the BICs, pseudo $R^2$ and estimated standard error, establishing which model is the best is difficult. What is easier is to eliminate the worst performing models. The “All” model at a grain size of 360 m radius was the easiest to eliminate, this had a much higher estimated standard error and, as well as including many more variables than the other models, it also had the highest BIC and lowest pseudo $R^2$. Next the 90 m combined and naio model had a BIC more than two
greater than the model with the lowest BIC. The 90 m “All” model has a BIC that is 1.91 greater than the best model. The model is just significant.

Every model contained MLGRI either under the mamane, naio canopy or under all trees. Some of the models only contained MLGRI, and were still as significant as models containing other variables. The coefficient is also negative as hypothesised, more heavily browsed sites transmitting more of the laser energy. MLGRI as a proxy for browsing appears to explain the palila count best.

Mamane have been shown in previous studies to produce less fruit during drought, and naio begin to die (Banko, 2014b; Banko et al., 2013). Some measure of water content in the foliage of the trees should have been a good indicator of available food for palila. Tree water content through water band index (WBI) was selected in only two models. When included in models with other variables, the coefficient was negative, however using WBI alone in a model of the 2009 bird count, the coefficient was almost always positive (being negative at the 90 m radius grain size for mamane), although with high error the coefficient could have been positive or negative. The negative coefficient seems more likely and is interesting as it was not what was hypothesised, with drier areas seeming to support more palila.

The assumption has been that the palila suffer during drought due to a reduction in the amount of fruit from mamane and their backup food supply naio (Banko et al., 2013). This negative coefficient for the combined WBI, suggests that in areas with more water there are fewer palila. Banko et al. (2013) showed that in drier years mamane produced on average 76% fewer pods. Therefore, the coefficient should be positive, with wetter areas supporting more palila. It is possible that this negative coefficient signifies the removal of the water limitation on black rats (Amarasekare, 1994). These wetter areas could be more susceptible to predation and therefore less appealing to palila. It is also possible that Mosquitoes are managing to breed in these wetter areas (Tucker-Mohl et al., 2010). Predation, disease, and increased food may act at different temporal and spatial scales,
an overall long-term trend and reduction in population, and a refuge from predation and disease in
drier areas at a shorter scale. These drier areas may have been wetter, but not too wet in the past
leading to more fruit and therefore more fecund palila or these areas may contain trees, which
quickly utilise water, leaving little standing water.

Tucker-Mohl et al. (2010) point out that despite low rainfall, malaria was present in the
lowland dry forest on the north-west side of the island of Hawaii. Tucker-Mohl et al. (2010)
discussed the work of LaPointe (2000) and the idea that feral pigs can create troughs in tree ferns
that hold water and allow mosquito to breed. Pigs are present in the mamane-naio woodland and are
not seen as an issue (Banko, 2014b); there are not any tree ferns however. It is possible that in these
wetter areas some other mechanism is allowing mosquitoes to breed. This should be investigated in
future work. Rats require additional water from that consumed in their food (Norman and
Baudinette, 1969). The amount of water obtained from their food will dictate how much extra is
needed. The WBI is measuring water in the vegetation, these areas therefore may not have
significant amounts of additional water, but the food of the rat contains more water itself. Future
research into water availability for rats could help distinguish if anything can be done to reduce
available water, which could only be achieved if the water is coming from pooled water and not
vegetation. Cho et al. (2010) used hyperspectral data to look at WBI in Eucalyptus grandis plantation
and showed that WBI along with other indicators could be used as indicators of site quality that had
been defined by total available water (TAW). Water content was also used by Jones et al. (2013) as
an indicator of bird species diversity. Despite these two uses, water content appears to be
underutilised.

Area of mamane was a significant predictor of the palila count at 180 and 360 m. The
coefficient was positive, suggesting that as area of mamane increases, the number of palila increase.
Even though area of naio and all trees was only found in one, model the negative coefficient is
consistent with there being less mamane where there are more naio. The coefficient being positive at larger scales for naio and all trees, also makes sense, at this scale it is important that naio exist as a secondary food source, even if they are not as useful at a small scale.

The mean patch size for mamane appeared in a 900 m radius grain size model, and mean patch size for naio was found in a 90 m radius grain size model. The coefficients were similar to that of the summed area, positive for mamane and negative for naio. The naio coefficient had the same inconsistency when looking at different scales, negative at lower scales and positive at higher. The high correlation between the summed area and the mean, suggests that where there are large areas of trees the mean patch size is high. It is therefore difficult to disentangle the patch size issue from an issue of available food. Larger patch size cannot be disproved by this analysis. The fact that, summed area or mean area is found at every scale adds further support for scale, potentially, being unimportant for overall modelling, but the switch in coefficient for naio suggests that mamane and naio may be being used at different scales. It also suggests that area in some form is second in importance to MLGRI; the two are not correlated to each other.

Tree height was included as a proxy for tree maturity. Tree height appears to be a less important predictor of the palila count. The naio height only appears in a marginally significant model, having the lowest p–value for the variable in that model or of any model, and mamane height only appears in insignificant models. Both have a positive coefficient. The positive coefficient is consistent with the hypothesised sign of the coefficient, taller more mature trees supporting more palila.

Elevation has been shown to have an effect on the predators, invertebrate pests, and invasive flora influencing palila. Elevation only appears in two insignificant models at the 90 m radius grain size. The coefficient is consistent in all cases with that hypothesised, with more palila found at higher elevation. WBI was also included at the 90 m grain size, it is possible that the water
content is the reason these pests are not found or found at high elevation. With elevation and WBI being imperfectly correlated, elevation may have appeared approximately correct in previous work, but water availability was the true reason for the apparent relationship with elevation.

Volume and tree count were not useful when predicting the palila count. In the case of the volume, it is likely that area is more powerful in predicting palila. Given that tree height was not powerful, it is possible due to the flat nature of the tree canopies that volume means less, potentially with not a huge increase in fruit for height and volume under the canopy. Likewise, an increase in the number of trees may not be as important as an increase in the total width of canopy, particularly if this study failed to capture the high number of stems under large areas of contiguous canopy.

MLGRI and area were the best two predictors of palila numbers, followed by WBI. MLGRI and area were both as hypothesised with less browsed areas and higher area of mamane supporting more palila. Dryer areas support more palila; this is the opposite relationship to what was hypothesised. A search of the literature would tend to suggest that despite Korpela et al. (2013) looking at the silhouette of branches using LiDAR intensity, this is the first time that this principle has been applied across a woodland or in relation to species habitat modelling to look at browsing. The next step would be to attempt to calibrate the measurement to a real quantitative measure and to confirm the observed pattern of browsing.

A buffer around each station was proposed to match the palila home range of 180 m radius derived from Farmer et al. (2014). This home range size did not agree with earlier research by Fancy et al. (1993), but appeared to give a clear consistent number that could be used. Other studies have used multiple scales in their analysis (Pedersen et al., 2013; Revermann et al., 2012). This analysis would tend to suggest that palila use multiple different scales, or different scales for different variables, potentially mamane at a small scale and naio at a larger scale. If being used for predicting
bird numbers from remotely sensed data, a multiscale analysis should be fitted to test data to allow for prediction.

This work could be improved through the collection of field data for training and testing. Training data could help in the classification of mamane and naio trees, allowing more advanced classification techniques to utilise more variables. For example, a decision tree may have allowed height to be used as a second level after LiDAR mean canopy intensity and NDVI, followed by other variables at lower levels. A testing dataset could also improve the attempts to validate the individual canopies as well as the classification. The lag in collection of remotely sensed data (2009) and analysis (2014) made the prospects of fieldwork less useful. Future work should attempt to plan fieldwork in conjunction with planned remote sensing. This would also allow for measures of water to be included in field sampling.

This study agrees with previous literature on many of the pressures effecting palila. It differs in that it picks out browsing, area of mamane trees and water content as the three most important. It suggests a means of illustrating browsing pressure in the mean LiDAR ground return intensity measure. It illustrates the importance of water in this dry forest giving an individual time frame perspective of water spatially across the site, as opposed to the scale of declines over multiple years. It also raises new questions. What causes mamane to fruit? Is water limiting the spread of black rats and mosquito into the mamane-naio forest? Future work to answer these questions could include mamane fruiting studies, possibly in a lab or experimental plot situation. Measures of water stress on the ground to quantify properly which trees are stressed and which are not. Taking blood samples of birds on the site similar to Tucker-Mohl et al. (2010) identifying the spatial extent of the spread of Malaria and trapping of rats to identify their spatial extent. Malaria and rat extents could then be compared to the field measurements of water stress and the water band index (WBI) from the hyperspectral data.
Chapter 6.

CONCLUSION

This study combined multiple variables obtained from remotely sensed data to investigate the interactions of palila, an endangered Hawaiian bird species with its remaining habitat on the southwest slopes of Mauna Kea. Remote sensing had not been used before on this site. Remote sensing allows for a view across the landscape to be taken. Previous research had looked at palila, this study re-examined some of the conclusions of this previous research, but also introduced new variables and means of quantifying habitat. Similar to previous research the size and distribution of both mamane and naio trees within the woodland, but this time continually across space using Light Detection and Ranging (LiDAR) data. The use of hyperspectral data allowed for the measurement of the amount of water in the canopy, through a ratio of two bands to calculate the water band index (WBI). Trends in rainfall across time had been examined in previous studies, however, the water distribution across the landscape in a single timeframe had not. LiDAR intensity was also utilised in this study. LiDAR intensity is often ignored or underutilised. This work utilised intensity in two ways, as means of classifying between mamane and naio tree species and of quantifying the density of the canopy, which is influenced by ungulate browsing.

Browsing is a major cause of the decline of palila. Some measure of this browsing was therefore needed. Browsing alters the branch size and canopy density of browsing trees. To quantify this reduction in branch size and canopy density, Mean LiDAR ground intensity (MLGRI) from below the trees was used. MLGRI showed that more heavily browsed areas gave higher penetration and reflection of intensity across the range of tree heights. The intensity measure performed exactly as hypothesised, both in its selection, but also in the nature of the relationship. Palila are more abundant in areas with a low ground return intensity, a dense canopy, therefore, indicating the negative effect that ungulate browsing has on so many aspects of palila habitat. The difference in
intensity is clearly observable either side of the boundary of the protected area, with more browsing taking place outside.

It had been hypothesised that the wetter areas would support a larger population of palila as more fruit would grow on mamane. WBI was selected for inclusion in the model, it however showed the opposite relationship, with dryer areas supporting more palila. It could have been that WBI was not working as a measure of wetness, somehow either failing to measure the canopy water content or the measure itself perform in the opposite way, with a different band relationship. An investigation of the WBI seemed to suggest that it was working. This therefore means that at least in the year the data was collected, 2009, that across the site the drier areas were supporting the higher populations of palila. This was not expected. Some other factor is therefore affecting palila in wet areas more negatively than the positive effect of more food being available. The most likely culprits are rats, although mosquitoes should also be considered. Mosquitoes and avian malaria have been shown to travel from breeding locations, suggesting that the effect of mosquitoes and malaria would both be found across the whole site and would unlikely be detected in a variable showing wetness of specific locations. These wetter areas could contain pools of water that are being drunk by rats to supplement the water rats obtain from their food. These wetter areas are also likely providing more water in the food the rats are eating. This makes rats the more likely culprit in this WBI relationship.

Mamane area within the buffer is consistent with canopy cover and mean patch size was consistent with woodland width, both identified in previous studies. Area and mean patch size (mean area), had the same relationship with the palila count, each appearing in different models. The two measures are highly correlated, although areas with higher area did not always match perfectly with mean patch size. To allow for differentiation between the two effects patch size would have to have higher variation across each different area. It is possible this could be achieved with many more survey location or maybe through some sort of audio bird survey. It is possible that the two
measures will always be linked, and that measuring both will never be useful. Some sort of bird selection modelling using radio tracking, may be the only way to assess patch size, looking to see which patch any individually tagged bird selects.

This study suffered from a lack of field data collected concurrently to the flying of the remotely sensed data in the year 2009. The delineation of individual trees and the classification of mamane and naio were not perfect. This is however far from simple uniform woodland. The mamane and naio are irregularly shaped, multi-stemmed, scattered trees. This work although imperfect managed to make a classification that did not rebel against the limited data that was available to investigate the quality of the estimated trees. Looking at individual canopies, in most circumstances, this analysis was successful at describing individual canopies. The classification of mamane and naio, again imperfect captured the main swaths of naio and mamane dominated areas. The use of LiDAR intensity was useful as the hyperspectral data could not differentiate between mamane and naio. This simple use of intensity however demonstrates the power of this underutilised component of LiDAR.

Buffers of 90, 180, 360 and 900 m were investigated in this study, all based on a home range size of 180 m (Farmer et al. 2014). Future work should include more buffer sizes to check the result that the aggregate grain size was unimportant, or to identify scales for inclusion in a multiscale model. Nighttime roosting information was not included in this study. Eventually a study combining information on nighttime and daytime habitat usage would be useful. As a first step, a study looking at nighttime roosting would be necessary for any future combined study.

Water band index should be ground truthed in future studies. How much water is needed by rats and mosquitoes? It might also be interesting to go back to previous work that included hyperspectral data and to see if water availability is important to the species or species being looked at in the study, although water is unlikely to be a limiting factor in many temperate studies. Pigs
should be investigated for their potential in creating standing water as a means of reversing the relationship of water across the landscape with palila.

The link between browsing and LiDAR intensity could also be investigated in studies that have already been conducted, reusing existing LiDAR data. Does this relationship only hold true for high-resolution LiDAR data, or could this be used to look at browsing across much greater scales? Data on browsing has been collected in the past during forest surveys. LiDAR intensity data could be used to look at browsing and to target culls or exclusions across whole regions or countries.

Water band index, should be further developed and tested and could provide new ways of looking at water across landscapes particularly in water limited environments. This study also showed the potential usefulness of remotely sensed water content in estimating rainfall. LiDAR intensity in this study, as a measure of browsing, is probably the clearest, most useful, transferable and revolutionary part of this study. Nobody has previously utilised LiDAR intensity in quantifying browsing. This measure can be used as a means of supporting the action that has been taken to remove browsers from palila habitat. It could also be used as a means of encouraging action outside the fenced protected area on Mauna Kea. This measure could have application worldwide where ungulates have been introduced or where they are running rampant due to the removal of population controlling predators.
REFERENCES


Banko, P.C., 2014b. May 2014 site visit.


77


Garcia-Feced, C., Tempel, D., Kelly, M., 2011. LiDAR as a tool to characterize wildlife habitat: California spotted owl nesting habitat as an example. J. For.


Henkin, D.L., 2009. PALILA (Loxioïdes bailleui) vs. Hawaii Department of Land and Natural Resources.


Moffiet, T., Mengersen, K., Witte, C., King, R., Denham, R., 2005. Airborne laser scanning: Exploratory data analysis indicates potential variables for classification of individual trees or


85
APPENDIX 1

Tree species classification R script

```r
rm(list=ls())

getwd()
setwd("G:\Work\MA\Palila\JC_Work\Spring2015\MarchStuff\FinalAnalysisFinal\ZonalCanopies")

listdbf <- list.files(pattern = "dbf")

library(foreign)

listdbf

for (i in 1:(length(listdbf)-4))
{
  if(i==1){
    DBF1 = read.dbf(listdbf[i], as.is = FALSE)
    Dataset = DBF1[,c("GRIDCODE2", "MEAN", "STD")]
    names(Dataset)[names(Dataset)="MEAN"] <- "MEAN1"
    names(Dataset)[names(Dataset)="STD"] <- "STD1"
    listdbf[i]
  } else {
    paste("MEAN", i)
    paste("STD",i)
    listdbf[i]
    Dataset[paste("MEAN",i, sep = "")]<- NA
    Dataset[paste("STD",i, sep = "")]<- NA
    DBF2 = read.dbf(listdbf[i], as.is = FALSE)
    Dataset[paste("MEAN",i, sep = "")]<- DBF2$MEAN * 1
    Dataset[paste("STD",i, sep = "")]<- DBF2$STD * 1
  }
}

test1 <- merge(read.dbf(listdbf[37], as.is = FALSE), read.dbf(listdbf[38], as.is = FALSE), by = "GRIDCODE2", all = TRUE)
test2 <- merge(read.dbf(listdbf[39], as.is = FALSE), read.dbf(listdbf[40], as.is = FALSE), by = "GRIDCODE2", all = TRUE)
names(test1)[names(test1)=="Gridcode2"] <- "GRIDCODE2"

names(test1)

test3 <- merge(test1,test2, by = "GRIDCODE2", all = TRUE)

Dataset2 = test3[,c("GRIDCODE2", "MEAN.x.x", "STD.x.x", "MEAN.y.x", "STD.y.x", "MEAN.x.y", "STD.x.y", "MEAN.y.y", "STD.y.y")]
names(Dataset2)[names(Dataset2)=="MEAN.x.x"] <- "MEAN37"
names(Dataset2)[names(Dataset2)=="STD.x.x"] <- "STD37"
names(Dataset2)[names(Dataset2)=="MEAN.y.x"] <- "MEAN38"
```

86
names(Dataset2)[names(Dataset2) == "STD.y.x"] <- "STD38"
names(Dataset2)[names(Dataset2) == "MEAN.x.y"] <- "MEAN39"
names(Dataset2)[names(Dataset2) == "STD.x.y"] <- "STD39"
names(Dataset2)[names(Dataset2) == "MEAN.y.y"] <- "MEAN40"
names(Dataset2)[names(Dataset2) == "STD.y.y"] <- "STD40"

Dataset3 <- merge(Dataset, Dataset2, by = "GRIDCODE2", all = TRUE)

# based on variable values
newdata <- na.omit(Dataset3)

remove(DBF1)
remove(DBF2)
remove(Dataset)
remove(Dataset2)
remove(test1)
remove(test2)
remove(test3)
remove(i)
remove(listdbf)

cols <- c(1,78,79,80,81)
DATA = newdata[,cols]

DATA <- na.omit(DATA)

Join = DATA[1]

for(i in 2:20)
{
  fit <- kmeans(DATA[2:5], i)
  aggregate(Join,by=list(fit$cluster),FUN=mean)
  Join <- data.frame(Join, fit$cluster)
}

write.dbf(Join, "G:\Work\MAI\Palila\JC_Work\Spring2015\MarchStuff\FinalAnalysisFinal\Classification\clustering_NDVIIntens2_20.dbf", factor2char = TRUE)

#ArcGIS Naio
APPENDIX 2
Regression analysis R script

rm(list=ls())

# New Data
getwd()
setwd("G:\Work\MA\Palila\JC_Work\Spring2015\MarchStuff\FinalAnalysisFinal\Analysis\Buffers\spatialJoin\TablesOnly\Consolodated")

if("XLConnect" %in% rownames(installed.packages()) == FALSE)
  {install.packages("XLConnect")}
library(XLConnect)               # load XLConnect package

wk = loadWorkbook("Buffer900m.xlsx")
df = readWorksheet(wk, sheet="Sheet1")

# Data for the birds
str_TreeSpreadsheetname <- "G:\Work\MA\Palila\JC_Work\Fall2014\Stations\Allvariables_hyper_area.txt"
stations <- read.table(str_TreeSpreadsheetname, sep="\t", header=TRUE)
BirdCounts = stations[c("StationID","X2008","X2009")]

BirdCounts$X2009
summary(BirdCounts$X2009)
var(BirdCounts$X2009)

names(df)[names(df)="Stations_i"] <- "StationID"

DATA <- merge(BirdCounts,df, by = "StationID", all = TRUE)

DATA["SQRElevMean"] <- DATA$MeanElev^2

names(DATA)
DATA <- na.omit(DATA)

names(DATA)
DATA$MeanWBI975All
DATA$MeanWBI975Ma
DATA$MeanWBI975Na

DATA["MeanWBI975All"] <- 1/DATA$MeanWBI975All
DATA["MeanWBI975Ma"] <- 1/DATA$MeanWBI975Ma
DATA["MeanWBI975Na"] <- 1/DATA$MeanWBI975Na

DATA$MeanWBI975Na[is.infinite(DATA$MeanWBI975Na)] <- 0
DATA$MeanWBI975Ma[is.infinite(DATA$MeanWBI975Ma)] <- 0
DATA$MeanWBI975All[is.infinite(DATA$MeanWBI975All)] <- 0

library(MASS)
library(boot)
# Mamane and Naio in a separate model

mamane <- glm.nb(X2009 ~ TreesMa + SumAreaPatchesMa + MeanPatchAreaMa + MeanHeightMa + SumVolumeMa + MeanGrnIntMa + MeanWBI975Ma + MeanElev, data=DATA)

summary(mamane)
mamane$R2 = mamane$deviance / mamane$null.deviance

mamane$LOOCV <- cv.glm(data = DATA, glmfit = mamane, K = nrow(DATA))

mamane$LOOCV$delta

naio <- glm.nb(X2009 ~ TreesNa + SumAreaPatchesNa + MeanPatchAreaNa + MeanHeightNa + SumVolumeNa + MeanGrnIntNa + MeanWBI975Na + MeanElev, data=DATA)

summary(naio)
naio$R2 = naio$deviance / naio$null.deviance

naio$LOOCV <- cv.glm(data = DATA, glmfit = naio, K = nrow(DATA))

naio$LOOCV$delta

mamane_stepAIC = stepAIC(mamane,k = log(nrow(DATA)))

summary(mamane_stepAIC)
mamane_stepAIC$R2 = mamane_stepAIC$deviance / mamane_stepAIC$null.deviance

mamane_stepAIC$LOOCV <- cv.glm(data = DATA, glmfit = mamane_stepAIC, K = nrow(DATA))

mamane_stepAIC$LOOCV$delta

confint(mamane_stepAIC)
confint.default(mamane_stepAIC)

naio_stepAIC = stepAIC(naio,k = log(nrow(DATA)))

summary(naio_stepAIC)
naio_stepAIC$R2 = naio_stepAIC$deviance / naio_stepAIC$null.deviance

naio_stepAIC$LOOCV <- cv.glm(data = DATA, glmfit = naio_stepAIC, K = nrow(DATA))

naio_stepAIC$LOOCV$delta

# Combined and separate

Separate <- glm.nb(X2009 ~ TreesMa + SumAreaPatchesMa + MeanPatchAreaMa + MeanHeightMa + SumVolumeMa + MeanGrnIntMa + MeanWBI975Ma + TreesNa + SumAreaPatchesNa + MeanPatchAreaNa + MeanHeightNa + SumVolumeNa + MeanGrnIntNa + MeanWBI975Na + MeanElev, data=DATA)

summary(Separate)
Separate$R2 = Separate$deviance / Separate$null.deviance

Separate$LOOCV <- cv.glm(data = DATA, glmfit = Separate, K = nrow(DATA))

Separate$LOOCV$delta

Combined <- glm.nb(X2009 ~ TreesAll + SumAreaPatchesAll + MeanPatchAreaAll + MeanHeightAll + SumVolumeAll + MeanGrnIntAll + MeanWBI975All + MeanElev, data=DATA)

summary(Combined)
Combined$R2 = Combined$deviance / Combined$null.deviance

Combined$LOOCV <- cv.glm(data = DATA, glmfit = Combined, K = nrow(DATA))

Combined$LOOCV$delta

Separate_stepAIC = stepAIC(Separate,k = log(nrow(DATA)))

summary(Separate_stepAIC)
Separate_stepAIC$R2 = Separate_stepAIC$deviance / Separate_stepAIC$null.deviance

Separate_stepAIC$R2
Separate_stepAIC$LOOCV <- cv.glm(data = DATA, glmfit = Separate_stepAIC, K = nrow(DATA))
Separate_stepAIC$LOOCV$delta

Combined_stepAIC = stepAIC(Combined,k = log(nrow(DATA)))
summary(Combined_stepAIC)
Combined_stepAIC$R2 = Combined_stepAIC$deviance / Combined_stepAIC$null.deviance
Combined_stepAIC$LOOCV <- cv.glm(data = DATA, glmfit = Combined_stepAIC, K = nrow(DATA))
Combined_stepAIC$LOOCV$delta

#All
All <- glm.nb(X2009 ~ TreesMa + SumAreaPatchesMa + MeanPatchAreaMa + MeanHeightMa +
SumVolumeMa + MeanGrnIntMa + MeanWB1975Ma + TreesNa + SumAreaPatchesNa + MeanPatchAreaNa +
MeanHeightNa + SumVolumeNa + MeanGrnIntNa + MeanWB1975Na + TreesAll + SumAreaPatchesAll +
MeanPatchAreaAll + MeanHeightAll + SumVolumeAll + MeanGrnIntAll + MeanWB1975All + MeanElev,
data=DATA)
summary(All)
All$R2 = All$deviance / All$null.deviance
All$LOOCV <- cv.glm(data = DATA, glmfit = All, K = nrow(DATA))
All$LOOCV$delta

All_stepAIC= stepAIC(All,k = log(nrow(DATA)))
summary(All_stepAIC)
All_stepAIC$R2 = All_stepAIC$deviance / All_stepAIC$null.deviance
All_stepAIC$LOOCV <- cv.glm(data = DATA, glmfit = All_stepAIC, K = nrow(DATA))
All_stepAIC$LOOCV$delta

#relative probability that the i-th model minimizes the (expected estimated) information loss
#probabilitymodel = exp((naio_stepAIC$aic - mamane_stepAIC$aic)/2)

#Prints for examination in R
{cat("Model \t AIC \t PsduoR2 \t EstPredErr EstPredErrAdj \n")
cat(sprintf("Elevation and sqrd model: \t %s \t %f \t %f \t %f\n", elevationQ$aic,
elevationQ$R2,elevationQ$LOOCV$delta[1], elevationQ$LOOCV$delta[2]))
cat(sprintf("elevation model: \t %s \t %f \t %f \t %f\n", elevation$aic,
elevation$R2,elevation$LOOCV$delta[1], elevation$LOOCV$delta[2]))
cat(sprintf("elevation sqrd model: \t %s \t %f \t %f \t %f\n", elevationSqr$aic,
elevationSqr$R2,elevationSqr$LOOCV$delta[1], elevationSqr$LOOCV$delta[2]))
cat(sprintf("Mamane model: \t %s \t %f \t %f \t %f\n", mamane$aic,
mamane$R2,mamane$LOOCV$delta[1], mamane$LOOCV$delta[2]))
cat(sprintf("Naio model: \t %s \t %f \t %f \t %f\n", naio$aic, naio$R2,naio$LOOCV$delta[1],
naio$LOOCV$delta[2]))
cat(sprintf("Mamane step model: \t %s \t %f \t %f \t %f\n", mamane_stepAIC$aic,
mamane_stepAIC$R2,mamane_stepAIC$LOOCV$delta[1], mamane_stepAIC$LOOCV$delta[2]))
cat(sprintf("Naio step model: \t %s \t %f \t %f \t %f\n", naio_stepAIC$aic,
naio_stepAIC$R2,naio_stepAIC$LOOCV$delta[1], naio_stepAIC$LOOCV$delta[2]))
cat(sprintf("Seperate model: \t %s \t %f \t %f \t %f\n", Separate$aic,
Separate$R2,Separate$LOOCV$delta[1], Separate$LOOCV$delta[2]))
cat(sprintf("Combined model: \t %s \t %f \t %f \t %f\n", Combined$aic,
Combined$R2,Combined$LOOCV$delta[1], Combined$LOOCV$delta[2])
}
cat(sprintf("Combined step model: \%s\%f\%s\%f\%s\%f\%s\%fn", Combined_stepAIC$aic, BIC(Combined_stepAIC), Combined_stepAIC$R2, Combined_stepAIC$LOOCV$delta[1], Combined_stepAIC$LOOCV$delta[2])))

#Prints better for transferring to excel
{cat("Model | AIC | BIC | PsduoR2 | EstPredErr | EstPredErrAdj")

cat(sprintf("All model: %f\%f\%f\%f\%f\%f\%fn", All$aic, All$R2, All$LOOCV$delta[1], All$LOOCV$delta[2])))

cat(sprintf("Naio model: %f\%f\%f\%f\%f\%f\%fn", naio$aic, BIC(naio), naio$R2, naio$LOOCV$delta[1], naio$LOOCV$delta[2])))

cat(sprintf("Mamane model: %f\%f\%f\%f\%f\%f\%fn", mamane$aic, BIC(mamane), mamane$R2, mamane$LOOCV$delta[1], mamane$LOOCV$delta[2])))

cat(sprintf("Seperate model: %f\%f\%f\%f\%f\%f\%fn", Seperate$aic, BIC(Seperate), Seperate$R2, Seperate$LOOCV$delta[1], Seperate$LOOCV$delta[2])))

cat(sprintf("Seperate step model: %f\%f\%f\%f\%f\%f\%fn", Seperate_stepAIC$aic, Seperate_stepAIC$R2, Seperate_stepAIC$LOOCV$delta[1], Seperate_stepAIC$LOOCV$delta[2])))

cat(sprintf("Combined step model: %f\%f\%f\%f\%f\%f\%fn", Combined_stepAIC$aic, Combined_stepAIC$R2, Combined_stepAIC$LOOCV$delta[1], Combined_stepAIC$LOOCV$delta[2])))

cat(sprintf("Mamane step model: %f\%f\%f\%f\%f\%f\%fn", mamane_stepAIC$aic, BIC(mamane_stepAIC), mamane_stepAIC$R2, mamane_stepAIC$LOOCV$delta[1], mamane_stepAIC$LOOCV$delta[2])))

cat(sprintf("Naio step model: %f\%f\%f\%f\%f\%f\%fn", naio_stepAIC$aic, BIC(naio_stepAIC), naio_stepAIC$R2, naio_stepAIC$LOOCV$delta[1], naio_stepAIC$LOOCV$delta[2])))

cat(sprintf("Seperate step model: %f\%f\%f\%f\%f\%f\%fn", Seperate_stepAIC$aic, BIC(Seperate_stepAIC), Seperate_stepAIC$R2, Seperate_stepAIC$LOOCV$delta[1], Seperate_stepAIC$LOOCV$delta[2])))

cat(sprintf("Naio step model: %f\%f\%f\%f\%f\%f\%fn", naio_stepAIC$aic, BIC(naio_stepAIC), naio_stepAIC$R2, naio_stepAIC$LOOCV$delta[1], naio_stepAIC$LOOCV$delta[2])))

cat(sprintf("Elevation and sqrd model: %f\%f\%f\%f\%f\%f\%fn", elevationQ$aic, BIC(elevationQ), elevation$R2, elevation$LOOCV$delta[1], elevation$LOOCV$delta[2])))

cat(sprintf("Seperate model: %f\%f\%f\%f\%f\%f\%fn", Seperate$aic, BIC(Seperate), Seperate$R2, Seperate$LOOCV$delta[1], Seperate$LOOCV$delta[2])))

cat(sprintf("Combined model: %f\%f\%f\%f\%f\%f\%fn", Combined$aic, BIC(Combined), Combined$R2, Combined$LOOCV$delta[1], Combined$LOOCV$delta[2])))

cat(sprintf("Elevation sqrd model: %f\%f\%f\%f\%f\%f\%fn", elevationSqr$aic, BIC(elevationSqr), elevationSqr$R2, elevationSqr$LOOCV$delta[1], elevationSqr$LOOCV$delta[2])))

cat(sprintf("Elevation model: %f\%f\%f\%f\%f\%f\%fn", elevation$aic, BIC(elevation), elevation$R2, elevation$LOOCV$delta[1], elevation$LOOCV$delta[2])))

cat(sprintf("Elevation and sqrd model: %f\%f\%f\%f\%f\%f\%fn", elevationQ$aic, BIC(elevationQ), elevation$R2, elevation$LOOCV$delta[1], elevation$LOOCV$delta[2])))

cat(sprintf("Mamane model: %f\%f\%f\%f\%f\%f\%fn", mamane$aic, BIC(mamane), mamane$R2, mamane$LOOCV$delta[1], mamane$LOOCV$delta[2])))

cat(sprintf("Naio model: %f\%f\%f\%f\%f\%f\%fn", naio$aic, BIC(naio), naio$R2, naio$LOOCV$delta[1], naio$LOOCV$delta[2])))

cat(sprintf("Mamane step model: %f\%f\%f\%f\%f\%f\%fn", mamane_stepAIC$aic, BIC(mamane_stepAIC), mamane_stepAIC$R2, mamane_stepAIC$LOOCV$delta[1], mamane_stepAIC$LOOCV$delta[2])))

cat(sprintf("Naio step model: %f\%f\%f\%f\%f\%f\%fn", naio_stepAIC$aic, BIC(naio_stepAIC), naio_stepAIC$R2, naio_stepAIC$LOOCV$delta[1], naio_stepAIC$LOOCV$delta[2])))

cat(sprintf("Seperate step model: %f\%f\%f\%f\%f\%f\%fn", Seperate_stepAIC$aic, BIC(Seperate_stepAIC), Seperate_stepAIC$R2, Seperate_stepAIC$LOOCV$delta[1], Seperate_stepAIC$LOOCV$delta[2])))

cat(sprintf("Combined model: %f\%f\%f\%f\%f\%f\%fn", Combined$aic, BIC(Combined), Combined$R2, Combined$LOOCV$delta[1], Combined$LOOCV$delta[2])))

cat(sprintf("Seperate model: %f\%f\%f\%f\%f\%f\%fn", Seperate$aic, BIC(Seperate), Seperate$R2, Seperate$LOOCV$delta[1], Seperate$LOOCV$delta[2])))

cat(sprintf("Combined step model: %f\%f\%f\%f\%f\%f\%fn", Combined_stepAIC$aic, Combined_stepAIC$R2, Combined_stepAIC$LOOCV$delta[1], Combined_stepAIC$LOOCV$delta[2])))

cat(sprintf("Combined step model: %f\%f\%f\%f\%f\%f\%fn", Combined_stepAIC$aic, BIC(Combined_stepAIC), Combined_stepAIC$R2, Combined_stepAIC$LOOCV$delta[1], Combined_stepAIC$LOOCV$delta[2])))

cat(sprintf("All model: %f\%f\%f\%f\%f\%f\%fn", All$aic, BIC(All), All$R2, All$LOOCV$delta[1], All$LOOCV$delta[2])))

cat(sprintf("All model: %f\%f\%f\%f\%f\%f\%fn", All$aic, BIC(All), All$R2, All$LOOCV$delta[1], All$LOOCV$delta[2])))

#Gives the coefficients for each model
{
  cat(length(coef(elevationQ)),"n")
  cat(length(coef(elevation)),"n")
  cat(length(coef(elevationSqr)),"n")
  cat(length(coef(mamane)),"n")
  cat(length(coef(naio)),"n")
  cat(length(coef(mamane_stepAIC)),"n")
  cat(length(coef(naio_stepAIC)),"n")
}
cat(length(coef(Seperate)),'n')
cat(length(coef(Combined)),'n')
cat(length(coef(Seperate_stepAIC)),'n')
cat(length(coef(Combined_stepAIC)),'n')
cat(length(coef(All)),'n')
cat(length(coef(All_stepAIC)),'n'))

# Each variable individually
IntensityMa <- glm.nb(X2009 ~ MeanGrnIntMa, data=DATA)
IntensityNa <- glm.nb(X2009 ~ MeanGrnIntNa, data=DATA)
IntensityCombined <- glm.nb(X2009 ~ MeanGrnIntAll, data=DATA)

IntensityMa$R2 = IntensityMa$deviance / IntensityMa$null.deviance
IntensityMa$LOOCV <- cv.glm(data = DATA, glmfit = IntensityMa, K = nrow(DATA))

IntensityNa$R2 = IntensityNa$deviance / IntensityNa$null.deviance
IntensityNa$LOOCV <- cv.glm(data = DATA, glmfit = IntensityNa, K = nrow(DATA))

IntensityCombined$R2 = IntensityCombined$deviance / IntensityCombined$null.deviance
IntensityCombined$LOOCV <- cv.glm(data = DATA, glmfit = IntensityCombined, K = nrow(DATA))

BIC(IntensityAll)
IntensityAll$R2
IntensityAll$LOOCV$delta[1]
summary(IntensityAll)

BIC(IntensityMa)
IntensityMa$R2
IntensityMa$LOOCV$delta[1]
summary(IntensityMa)

IntensityMa
IntensityNa
IntensityCombined

BIC(IntensityNa)
IntensityNa$R2
IntensityNa$LOOCV$delta[1]

BIC(IntensityCombined)
IntensityCombined$R2
IntensityCombined$LOOCV$delta[1]

IntensityAllstepBIC = stepAIC(IntensityAll,k = log(nrow(DATA)))

summary(IntensityAllstepBIC)

IntensityAllstepAIC = stepAIC(IntensityAll)
summary(IntensityAllstepAIC)

Water <- glm.nb(X2009 ~ MeanWBI975Ma + MeanWBI975Na + MeanWBI975All, data=DATA)

WaterMa <- glm.nb(X2009 ~ MeanWBI975Ma, data=DATA)
WaterMa$R2 = WaterMa$deviance / WaterMa$null.deviance
WaterMa$LOOCV <- cv.glm(data = DATA, glmfit = WaterMa, K = nrow(DATA))
WaterMa$R2
WaterMa$LOOCV$delta[1]

WaterNa <- glm.nb(X2009 ~ MeanWBI975Na, data=DATA)
WaterNa$R2 = WaterNa$deviance / WaterNa$null.deviance
WaterNa$LOOCV <- cv.glm(data = DATA, glmfit = WaterNa, K = nrow(DATA))
WaterNa$R2
WaterNa$LOOCV$delta[1]

WaterAll <- glm.nb(X2009 ~ MeanWBI975All, data=DATA)
WaterAll$R2 = WaterAll$deviance / WaterAll$null.deviance
WaterAll$LOOCV <- cv.glm(data = DATA, glmfit = WaterAll, K = nrow(DATA))

WaterMa$R2
WaterMa$LOOCV$delta[1]
coeff(WaterMa)
summary(WaterMa)
plot(WaterMa)

WaterNa$R2
WaterNa$LOOCV$delta[1]
coeff(WaterNa)
WaterAll$R2
WaterAll$LOOCV$delta[1]
coeff(WaterAll)

plot(DATA$MeanWBI975Ma, DATA$X2009)

TreesMa + SumAreaPatchesMa + MeanPatchAreaMa + MeanHeightMa + SumVolumeMa +
MeanGrnIntMa + MeanWBI975Ma + TreesNa + SumAreaPatchesNa + MeanPatchAreaNa + MeanHeightNa
+ SumVolumeNa + MeanGrnIntNa + MeanWBI975Na + TreesAll + SumAreaPatchesAll +
MeanPatchAreaAll + MeanHeightAll + SumVolumeAll + MeanGrnIntAll + MeanWBI975All + MeanElev

SumAreaMa <- glm.nb(X2009 ~ SumAreaPatchesMa, data=DATA)
SumAreaMa$R2 = SumAreaMa$deviance / SumAreaMa$null.deviance
SumAreaMa$LOOCV <- cv.glm(data = DATA, glmfit = SumAreaMa, K = nrow(DATA))
SumAreaMa$R2
SumAreaMa$LOOCV$delta[1]

SumAreaNa <- glm.nb(X2009 ~ SumAreaPatchesNa, data=DATA)
SumAreaNa$R2 = SumAreaNa$deviance / SumAreaNa$null.deviance
SumAreaNa$LOOCV <- cv.glm(data = DATA, glmfit = SumAreaNa, K = nrow(DATA))
SumAreaNa$R2
SumAreaNa$LOOCV$delta[1]

SumAreaAll <- glm.nb(X2009 ~ SumAreaPatchesAll, data=DATA)
SumAreaAll$R2 = SumAreaAll$deviance / SumAreaAll$null.deviance
SumAreaAll$LOOCV <- cv.glm(data = DATA, glmfit = SumAreaAll, K = nrow(DATA))

SumAreaMa$R2
SumAreaMa$LOOCV$delta[1]
coeff(SumAreaMa)
SumAreaNa$R2
SumAreaNa$LOOCV$delta[1]
coeff(SumAreaNa)
SumAreaAll$R2
SumAreaAll$LOOCV$delta[1]
MeanAreaMa <- glm.nb(X2009 ~ MeanPatchAreaMa, data=DATA)
MeanAreaMa$R2 = MeanAreaMa$deviance / MeanAreaMa$null.deviance
MeanAreaMa$LOOCV <- cv.glm(data = DATA, glmfit = MeanAreaMa, K = nrow(DATA))
MeanAreaMa$R2
MeanAreaMa$LOOCV$delta[1]

MeanAreaNa <- glm.nb(X2009 ~ MeanPatchAreaNa, data=DATA)
MeanAreaNa$R2 = MeanAreaNa$deviance / MeanAreaNa$null.deviance
MeanAreaNa$LOOCV <- cv.glm(data = DATA, glmfit = MeanAreaNa, K = nrow(DATA))
MeanAreaNa$R2
MeanAreaNa$LOOCV$delta[1]

MeanAreaAll <- glm.nb(X2009 ~ MeanPatchArea, data=DATA)
MeanAreaAll$R2 = MeanAreaAll$deviance / MeanAreaAll$null.deviance
MeanAreaAll$LOOCV <- cv.glm(data = DATA, glmfit = MeanAreaAll, K = nrow(DATA))
MeanAreaAll$R2

HeightMa <- glm.nb(X2009 ~ MeanHeightMa, data=DATA)
HeightMa$R2 = HeightMa$deviance / HeightMa$null.deviance
HeightMa$LOOCV <- cv.glm(data = DATA, glmfit = HeightMa, K = nrow(DATA))
HeightMa$R2
HeightMa$LOOCV$delta[1]

HeightNa <- glm.nb(X2009 ~ MeanHeightNa, data=DATA)
HeightNa$R2 = HeightNa$deviance / HeightNa$null.deviance
HeightNa$LOOCV <- cv.glm(data = DATA, glmfit = HeightNa, K = nrow(DATA))
HeightNa$R2
HeightNa$LOOCV$delta[1]

HeightAll <- glm.nb(X2009 ~ MeanHeightAll, data=DATA)
HeightAll$R2 = HeightAll$deviance / HeightAll$null.deviance
HeightAll$LOOCV <- cv.glm(data = DATA, glmfit = HeightAll, K = nrow(DATA))
HeightMa$R2
HeightMa$LOOCV$delta[1]

HeightMa <- glm.nb(X2009 ~ MeanHeightMa, data=DATA)
HeightMa$R2 = HeightMa$deviance / HeightMa$null.deviance
HeightMa$LOOCV <- cv.glm(data = DATA, glmfit = HeightMa, K = nrow(DATA))
HeightMa$R2
HeightMa$LOOCV$delta[1]

HeightNa <- glm.nb(X2009 ~ MeanHeightNa, data=DATA)
HeightNa$R2 = HeightNa$deviance / HeightNa$null.deviance
HeightNa$LOOCV <- cv.glm(data = DATA, glmfit = HeightNa, K = nrow(DATA))
HeightNa$R2
HeightNa$LOOCV$delta[1]

HeightAll <- glm.nb(X2009 ~ MeanHeightAll, data=DATA)
HeightAll$R2 = HeightAll$deviance / HeightAll$null.deviance
HeightAll$LOOCV <- cv.glm(data = DATA, glmfit = HeightAll, K = nrow(DATA))
HeightMa$R2
HeightMa$LOOCV$delta[1]

HeightMa <- glm.nb(X2009 ~ MeanHeightMa, data=DATA)
HeightMa$R2 = HeightMa$deviance / HeightMa$null.deviance
HeightMa$LOOCV <- cv.glm(data = DATA, glmfit = HeightMa, K = nrow(DATA))
HeightMa$R2
HeightMa$LOOCV$delta[1]

HeightNa <- glm.nb(X2009 ~ MeanHeightNa, data=DATA)
HeightNa$R2 = HeightNa$deviance / HeightNa$null.deviance
HeightNa$LOOCV <- cv.glm(data = DATA, glmfit = HeightNa, K = nrow(DATA))
HeightNa$R2
HeightNa$LOOCV$delta[1]

HeightAll <- glm.nb(X2009 ~ MeanHeightAll, data=DATA)
HeightAll$R2 = HeightAll$deviance / HeightAll$null.deviance
HeightAll$LOOCV <- cv.glm(data = DATA, glmfit = HeightAll, K = nrow(DATA))
HeightMa$R2
HeightMa$LOOCV$delta[1]

HeightMa$R2
HeightMa$LOOCV$delta[1]

ElevAll <- glm.nb(X2009 ~ MeanElev, data=DATA)
ElevAll$R2 = ElevAll$deviance / ElevAll$null.deviance

94
ElevAll$LOOCV <- cv.glm(data = DATA, glmfit = ElevAll, K = nrow(DATA))

ElevAll$R2
ElevAll$LOOCV$delta[1]
coef(ElevAll)

VolMa <- glm.nb(X2009 ~ MeanWBI975Ma, data=DATA)
VolMa$R2 = VolMa$deviance / VolMa$null.deviance
VolMa$LOOCV <- cv.glm(data = DATA, glmfit = VolMa, K = nrow(DATA))
VolMa$R2
VolMa$LOOCV$delta[1]

coef(IntensityMa)[2]
coef(IntensityNa)[2]
coef(IntensityCombined)[2]
APPENDIX 3
Visual tree comparison

2000 survey count 10
Delineated count 9

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

7                              5

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

8

2

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

17

Delineated count

3

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m$^2$)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

Station

Buffer

Delineated canopies

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

17                              2

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

16

22.6 m radius (circle area 1600 m²)

Delineated count

1
2000 survey count: 15
Delineated count: 3

22.6 m radius (circle area 1600 m²)
2000 survey count

15

Delineated count

15

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Station

Buffer

Delineated canopies

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count 89
Delineated count 31

22.6 m radius (circle area 1600 m²)
2000 survey count 31  Delineated count 28

Station

Buffer

Delineated canopies

22.6 m radius (circle area 1600 m²)
2000 survey count
54

Delineated count
42

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count
Delineated count

96                              28

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Delineated count

30

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

D

Station

Buffer

Delineated canopies

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

56

Delineated count

29

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count
150

Delineated count
36

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Delineated count

88                              24

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

52                              39

22.6 m radius (circle area 1600 m$^2$)

Station

Buffer

Delineated canopies
2000 survey count: 138
Delineated count: 44

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

66                              31

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count count

36                              21

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

82

Delineated count

27

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

76

Delineated count

28

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count: 44
Delineated count: 18

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

43                              21

22.6 m radius (circle area 1600 m$^2$)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Station

Buffer

Delineated canopies

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

Delineated canopies

22.6 m radius (circle area 1600 m²)
2000 survey count

32

Delineated count

19

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count: 63
Delineated count: 34

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

46                              44

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

78                              46

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
Delineated count

22.6 m radius (circle area 1600 m²)

2000 survey count

74

50
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
36                              18
2000 survey count
Delineated count

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count: 23
Delineated count: 12

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

38                              13

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count: 29
Delineated count: 11

22.6 m radius (circle area 1600 m²)

- Station
- Buffer
- Delineated canopies
2000 survey count

77

Delineated count

42

22.6 m radius (circle area 1600 m²)
2000 survey count

35

Delineated count

25

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

72                              35

22.6 m radius (circle area 1600 m$^2$)

Station

Buffer

Delineated canopies
2000 survey count

58

Delineated count

18

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count: 58
Delineated count: 39

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Delineated count

Delineated canopies

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22                              12

Delineated canopies

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count: 66

Delineated count: 19

22.6 m radius (circle area 1600 m²)
2000 survey count 16  Delineated count 11

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

35                              32

Delineated count

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count: 14, Delineated count: 7

22.6 m radius (circle area 1600 m²)

Station, Buffer, Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m$^2$)

Station

Buffer

Delineated canopies
Delineated count

22.6 m radius (circle area 1600 m²)

2000 survey count

Delineated canopies

Station

Buffer
2000 survey count

Delineated count

Station

Buffer

Delineated canopies

22.6 m radius (circle area 1600 m²)
2000 survey count

27

Delineated count

17

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count: 14
Delineated count: 9

22.6 m radius (circle area 1600 m²)

- Station
- Buffer
- Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

30                              16

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count: 17
Delineated count: 19

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Station

Buffer

Delineated canopies

Delineated count

22.6 m radius (circle area 1600 m²)

16

7
2000 survey count

Delineated count

27                              23

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m$^2$)
2000 survey count

Delineated count

16 7

22.6 m radius (circle area 1600 m$^2$)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

27

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

44

Delineated count

21

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

43                              27

Delineated count

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

38                              35

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

47                              37

22.6 m radius (circle area 1600 m²)
2000 survey count: 58
Delineated count: 47

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)
2000 survey count

Delineated count

67                              40

22.6 m radius (circle area 1600 m²)

Station

Buffer

Delineated canopies
2000 survey count

Delineated count

22.6 m radius (circle area 1600 m²)

Station
Buffer
Delineated canopies