

**USE OF REMOTE SENSING TO GENERATE A PREDICTIVE MODEL OF
BOVINE TUBERCULOSIS**

DANIEL JAMES RUSSELL

A dissertation presented in partial fulfilment of the requirements of the Masters of
Veterinary Studies

MASSEY UNIVERSITY

2004

ACKNOWLEDGEMENTS

I take my hat off to anyone who undertakes postgraduate study while maintaining full-time employment. I have found it to be a long, but extremely rewarding road to travel down. I am very grateful to the EpiCentre for providing a very supportive environment to work in.

I would like to thank my two supervisors, Dr. Jo McKenzie and Dr. Nigel Perkins. Nigel has been an amazing role model and is always willing to share his wealth of knowledge with me – I am truly indebted. I am amazed at how he can tackle any problem with such ease. Jo is also a top-notch supervisor who added the epidemiology of Tb knowledge to the study and was always willing to help throughout the study. Both supervisors have been extremely positive and extremely patient!

To the students of the EpiCentre – thank you for all the fun times at Icons, and for sprinkling glitter all through my car.

A special thanks to Mel and my family who have been as supportive as usual.

TABLE OF CONTENTS

CHAPTER 1: LITERATURE REVIEW	7
1. PART 1. Introduction to Remote Sensing.....	7
1.1. Introduction	7
1.2. Data Acquisition.....	7
1.3. Sensors and Instruments.....	8
1.4. Visible and Near-Infrared Sensors	9
1.4.1. Thermal	9
1.4.2. Microwave.....	9
1.5. Data Resolution and Extent.....	10
1.5.1. Spectral.....	10
1.5.2. Spatial.....	10
1.5.3. Temporal	10
1.5.4. Radiometric	11
1.6. Satellites	11
2. PART 2. APPLICATIONS OF REMOTE SENSING TO EPIDEMIOLOGY	12
2.1. Vegetation Classification	12
2.1.1. Classification Methods.....	12
2.1.1.1. Unsupervised Classification.....	12
2.1.1.2. Supervised Classification	13
2.1.2. Accuracy of Classification	13
2.1.3. Application of Vegetation Classification	14
2.2. Habitat Patterns	15
2.2.1. Landscape Indices	16
2.2.1.1. Contagion	16
2.2.1.2. Mean Nearest-Neighbour Distance	17
2.2.1.3. Mean Patch Fractal Dimension	17
2.2.1.4. Edge Density	17
2.2.1.5. Mean Proximity Distance.....	17
2.2.1.6. Core Area Density.....	18
2.2.2. Scale	18
2.2.3. Applications of Landscape Indices	18

2.3.	Vegetation Indices	19
2.3.1.	Description of Vegetation Indices	20
2.3.1.1.	Ratio Vegetation Index (RVI)	20
2.3.1.2.	Normalised Difference Vegetation Index (NDVI)	20
2.3.1.3.	PVI.....	20
2.3.1.4.	Soil-Adjusted Vegetation Index (SAVI)	20
2.3.2.	Factors Affecting Vegetation Indices	21
2.3.3.	Application of Vegetation Indices	21
2.4.	Vegetation Classes Versus Indices	23
2.5.	Other Environmental Variables	23
3.	Conclusion	25
4.	Reference List.....	25
CHAPTER 2: USE OF REMOTE SENSING TO CREATE A PREDICTIVE MODEL OF BOVINE TUBERCULOSIS		31
1.	Introduction.....	31
2.	Materials and Methods	32
2.1.	Vegetation Image.....	32
2.2.	Calculation of Habitat Variables	32
2.3.	Selection of Farms	33
2.4.	Dependent Variable	34
2.5.	Statistical Analysis.....	34
3.	Results.....	35
4.	Discussion.....	39
5.	Acknowledgements.....	41
6.	References.....	41
APPENDICES		43
1.	Appendix A: Principle Components Analysis.....	43
1.1.	Podocarp / Broadleaf	43
1.2.	Broadleaf Scrub / Pure.....	44
1.3.	Broadleaf Scrub / Manuka	45
1.4.	Pine / Mixed.....	46
1.5.	Willows.....	47
1.6.	Manuka / Gorse.....	48
1.7.	Pine	49

1.8.	Manuka.....	50
1.9.	Gorse	51
2.	Appendix B: Individual Class Models	53
2.1.	Prerequisites	53
2.2.	Podocarp / Broadleaf.....	54
2.2.1.	SAS Code	54
2.2.2.	Model Result	54
2.2.3.	Model Check	55
2.3.	Broadleaf Scrub / Pure	55
2.3.1.	SAS Code	55
2.3.2.	Model Result	56
2.3.3.	Model Check	56
2.4.	Broadleaf Scrub / Manuka.....	56
2.4.1.	SAS Code	56
2.4.2.	Model Result	57
2.4.3.	Model Check	57
2.5.	Pine / Mixed	57
2.5.1.	SAS Code	57
2.5.2.	Model Result	58
2.5.3.	Model Check	58
2.6.	Willows	58
2.6.1.	SAS Code	58
2.6.2.	Model Result	59
2.6.3.	Model Check	59
2.7.	Manuka / Gorse	60
2.7.1.	SAS Code	60
2.7.2.	Model Result	60
2.7.3.	Model Check	60
2.8.	Pine.....	61
2.8.1.	SAS Code	61
2.8.2.	Model Result	61
2.8.3.	Model Check	62

2.9.	Manuka	62
2.9.1.	SAS Code.....	62
2.9.2.	Model Result.....	62
2.9.3.	Model Check.....	63
2.10.	Gorse.....	63
2.10.1.	SAS Code.....	63
2.10.2.	Model Result.....	63
2.10.3.	Model Check.....	64

CHAPTER 1: LITERATURE REVIEW

1. PART 1. Introduction to Remote Sensing

1.1. Introduction

Harrison and Jupp (1990) define remote sensing as "a general term which describes the action of obtaining information about an object with a sensor which is physically separated from the object". The first uses of remote sensing date back to the early 19th century with the invention of photography. The idea of taking a photo of the earth's surface from an elevated platform occurred in 1859 when a balloonist took a photo of Petit Bicêtre near Paris, and the first photographs taken of the earth from space were released in the early 1960's (Harrison and Jupp 1990). Since this time, there have been rapid developments in remote sensing technology (mostly due to military applications). As the technology shifted from military to civilian domains, many different scientific areas have found remote sensing beneficial, see Table 1.

The two main platforms used in terrestrial remote sensing are aircraft and satellites. Aircraft have the advantage of flying at low altitudes and therefore recording observations at a fine scale whereas satellites have the advantage of regular and large scale coverage of an area. Depending on the type of image required, cost is both an advantage and disadvantage of each platform. There are many ways to gather remotely sensed information e.g. photography, satellite images of earth, human eye, sonar etc. The primary focus of this section is to introduce remote sensing with emphasis on data acquisition about the earth's surface by remote sensors onboard satellites.

Table 1 Examples where terrestrial remote sensing has been applied.

Atmospheric pollution	Meteorology	Cartology
Chemical pollution	Climatology	Geography
Agricultural crop health	Natural disasters	Ecology
Oceanography	Archaeology	Epidemiology

Adapted from Cracknell and Hayes 1991.

1.2. Data Acquisition

Remote sensing devices are designed to detect various types of energies, such as electromagnetic (EM), gravity, magnetism and sound waves. EM is the most commonly used by earth remote sensing devices (Harrison and Jupp 1990). The electromagnetic spectrum, of which visible light

makes up a very small proportion, consists of EM Radiation (EMR) at differing wavelengths. EMR is continuously emitted from all objects that are above absolute zero temperature (-273.15C). The adsorption of EMR by an object is dependant on both temperature and type of material, with 'black' material absorbing all EMR that it receives. The way in which EMR is reflected by an object is affected by the smoothness of the object's surface relative to the wavelength. Surfaces that cause scattering of EMR in all directions are referred to as diffuse or lambertian. Some regions of the EM spectrum are completely absorbed by gases within the atmosphere and therefore cannot be used by remote sensors. Furthermore, EMR that does pass through the earth's atmosphere is still subject to scattering by fog, haze and cloud etc. The various types of scattering reduce the contrast in an image, although image processing techniques are available to eliminate or reduce these atmospheric 'noises'.

Remote sensing works on the same principles to vision, in that visual interpretation is achieved by the type of EMR that is reflected from the object. However, a major advantage of remote sensors is that it can 'see' reflectance of EMR that is not in the visible spectrum. Remote sensors also have the ability to record EMR that they receive in electronic format. Radiation received by the sensor can be converted into a grid image. For each grid, the average radiance received by the scanner is converted into a Digital Number (DN), which are positive integers in 8-, 9- and 10-bit binary computer coding scales (i.e. 2^8 , 2^9 and 2^{10}). These DN's enable the image to be readily analysed with a computer.

1.3. Sensors and Instruments

Sensors are the devices that detect EMR. Remote sensing systems are either active or passive. In an active system (e.g. radars), the remote sensing instrument emits a signal itself towards a target and then detects the returned signal. Passive systems do not emit a signal, but simply detect any radiation signals that happen to arrive at the sensor. Therefore, passive systems are restricted to detecting radiation that is emitted with reasonable intensity and quantity from the sun and reflected by the earth whereas active systems are restricted to wavelengths in which a reasonable intensity can be generated by the remote device itself.

Remote sensing instruments can also be divided into imaging and non-imaging devices. Imaging devices create a two-dimensional image of a surface (e.g. earth or clouds) and may detect variations in colour, temperature and surface topology. Non-imaging devices acquire information such as the height of the satellite above the earth or surface roughness. Data acquired from imaging devices has

a higher processing demand, such as adjusting for atmospheric effects, before the image is ready for the end user. The type of sensor is also usually classified by the type(s) of EMR that it can detect and sensors may detect radiation at one or many areas of the EM Spectrum. These areas are referred to as 'channels' or 'bands'.

1.4. Visible and Near-Infrared Sensors

Multi-spectral scanners (MSS) detect radiation at multiple wavelengths in the visible, near infrared, middle infrared and thermal infrared parts of the EM spectrum. MSS are able to take advantage of the fact that surfaces reflect radiation differently in different parts of the spectrum. The spectrum of wavelengths reflected by a surface is sometimes referred to as the “spectral signature” of an object. EMR in these regions of the spectrum are affected by atmospheric scattering and therefore these devices are limited by atmospheric conditions. The near-infrared wavelengths are particularly important in forestry, agriculture, epidemiological and ecological studies, as vegetation reflects strongly in the near-infrared parts of the spectrum. Most MSS in the visible and near infrared regions are passive. Lidar (Laser Imaging raDAR) is an example of an active sensor in the ultraviolet to near infrared regions which requires clear atmospheric conditions, and as a result these are usually aircraft-borne. Lidar sensors have been used in mapping, air and pollution studies (Harrison and Jupp 1990).

1.4.1. Thermal

Thermal sensors are a passive system. Sensors that detect wavelengths in the thermal infrared area of the spectrum allow measurement of relative temperatures of surfaces. Actual temperatures can be calculated for surfaces that have a known emissivity (e.g. water/sea), but the interpretation of thermal-infrared over land areas is difficult. Thermal-infrared observations are also restricted to areas of no cloud.

1.4.2. Microwave

Microwave sensors are either passive or active. Microwave remote sensing can be used in almost all weather conditions except heavy rain. The intensity of microwave radiation emitted or reflected from the earth's surface is very small, which means passive remote sensing systems need sensitive

sensors or must obtain information over a large area (i.e. low spatial resolution). However, active systems are advantaged by the lack of background 'noise'.

1.5. Data Resolution and Extent

Data resolution refers to the intensity or rate of sampling, whereas extent refers to the overall coverage of the dataset. These two characteristics determine the ability to differentiate between features of interest in a digital image and hence determine the suitability of a remotely sensed image to the research objectives. Data resolution and extent can be described as having four properties: spectral, spatial, temporal and radiometric. Each of these properties is discussed below.

1.5.1. Spectral Resolution

Reflection of radiation from an object is a continuous function of wavelength, but remote sensing systems are only able to measure reflectance in discrete bands (channels). Spectral resolution is related to the number and width of these channels. Spectral extent refers to the range of wavelengths being sensed from all bands. Higher spectral resolution allows greater discrimination between objects on the surface of the earth. This has been particularly important in epidemiological studies where classification of different vegetation types is required.

1.5.2. Spatial Resolution

Spatial resolution refers to the smallness of objects within an image that can be distinguished as separate entities (Harrison and Jupp 1990). Therefore spatial resolution of an image is related to pixel size and for satellite images, spatial resolution is usually described in terms of the area of ground that each pixel covers (e.g. 20 by 20 m). Spatial extent refers to the overall spatial coverage of the image. Spatial resolution of current sensors range from low ($>2\text{km}^2$) to high (0.6m^2). Spatial resolution affects how we perceive the heterogeneity of the landscape. A heterogeneous landscape measured at an increasingly coarser scale will tend towards a homogeneous landscape (Pickett and Cadenasso 1995).

1.5.3. Temporal Resolution

Temporal resolution refers to the shortest reliable return period of the satellite over a particular area (Thomson and Conner 2000). Temporal extent refers to the total period over which imagery is available (Harrison and Jupp 1990). There is usually a trade-off between spatial and temporal

resolution. Satellites that provide high temporal resolution usually have images of low spatial resolution.

1.5.4. Radiometric Resolution

Radiometric resolution determines how finely the sensor can discriminate between objects of similar reflection. It is inversely proportional to the contrast of an image. Sensors that have fine radiometric resolution are more sensitive at detecting a small difference in reflected or emitted radiation.

1.6. Satellites

Satellites are divided into two main types: polar-orbiting earth observation and geostationary meteorological satellites. Geostationary satellites are positioned at a high altitude above the earth's surface (~36000 Km) and have the same angular velocity to the earth's rotation. Therefore, these satellites are always observing the same portion of the earth's surface and produce relatively low spatial resolution images. The principle sensor on these satellites is the Visible and Infrared Spin Scan Radiometer (VISSR).

Polar-orbiting satellites are positioned at various altitudes (700-1500Km), providing near global coverage with images at various resolutions. Data from these satellites may be sourced on request. These satellites have a relatively small spatial extent (60 - 176 km²) and can provide high spatial resolution images (e.g. 20m²) which makes them most suited to small area studies (Thomson and Conner 2000).

It is anticipated that about 80 earth orbiting missions, with more than 200 sensor instruments onboard, will take place in the next 15 years (Beck *et al* 2000). These satellites will have improved spatial, spectral and temporal resolution, and operate at a lower cost. Beck (2000) illustrated current and future satellites that will be useful for the mapping of soil moisture, urban features and vegetation type. Some of the future satellites will have spatial and temporal resolutions between 1-5 m² and 2-7 days respectively. This advanced technology will provide more environmental variables for studying the epidemiology of disease.

2. PART 2. APPLICATIONS OF REMOTE SENSING TO EPIDEMIOLOGY

The usefulness of remote sensing (RS) in epidemiology is related to the correlation between mapped environmental variables and disease dynamics and distributions (Boone *et al* 2000). One of the most common uses of remotely sensed data is to derive vegetation or habitat data. These variables can be grouped into three major classes: vegetation classes, habitat patterns, and vegetation indices.

2.1. Vegetation Classification

It has been useful to classify different vegetation types as the distribution of disease reservoir and host may relate to the occurrence and distribution of specific vegetation types rather than broad vegetation cover (Beck *et al* 2000). Different vegetation types may also be useful indicators of environmental characteristics, including moisture, soil type, and elevation (Boone *et al* 2000). Certain plant communities also affect the distribution of parasites, such as ticks, through modification of microclimate of living conditions of their hosts providing favourable food resources and shelters (Daniel *et al* 1998; Mills and Childs 1998). Vegetation classification involves creating a digital image that identifies specific vegetation types. These vegetation types can either be homogeneous in composition e.g. pine trees or grassland or heterogeneous e.g. a mixture of gorse and Manuka.

2.1.1. Classification Methods

The two methods of classifying vegetation into different classes are supervised and non-supervised. The choice of method depends on whether or not the researcher can identify known areas of vegetation classes. It is possible to use both methods to classify vegetation, for example, Daniel *et al* (1998) used both a supervised classification followed by non-supervised classification to achieve a deeper recognition of the vegetation structure.

2.1.1.1. Unsupervised Classification

Unsupervised classification involves classifying vegetation in a poorly known area, using only mathematical algorithms. Clustering algorithms group together pixels with similar spectral properties then assigns them to a class. This may be useful to epidemiologists when land cover classes affecting disease/vector are not known, however it is not guaranteed that these classes will be of epidemiological significance (Curran *et al* 2000).

2.1.1.2. Supervised Classification

Supervised classification involves incorporating prior knowledge about the study area before clustering similar spectral signatures into vegetation classes. This approach has been useful for epidemiologists because land cover classes that affect disease/vectors are often known (Curran *et al* 2000).

The supervised process involves three stages. Firstly, the 'training stage' involves matching areas of known land cover with the image's spectral signature such that each land cover class can be described by some spectral statistics. These spectral statistics are then used, in conjunction with an algorithm, to match each pixel to the most appropriate land cover class. The algorithms used are either statistical or non-statistical in their approach. A description of these different algorithms is beyond this review, but it is important to realise that each type of algorithm varies in its ability to correctly classify land cover. The final stage involves testing the accuracy of the classification. This is usually displayed as a cross-tabulation of actual versus predicted classes, using areas of known land cover that were not used in the training stage.

2.1.2. Accuracy of Classification

The success of supervised vegetation classification is reliant on the classes being well discriminated in the spectral space (Andrefouet and Claereboudt 2000). Misclassification occurs when a pixel is assigned to an incorrect vegetation class. Accuracy can be thought of as the level of agreement between the actual ground measurements and the classified vegetation image as neither the ground data or the classification process is 100% accurate. Table 2 gives a brief description of accuracy measurements.

Table 2 Measurements that describe the accuracy of land cover classification

Measure	Description
Overall accuracy %	The count of the correctly allocated cases divided by all cases used in the accuracy test.
User's accuracy %	The count of the correct actual class cases divided by all cases in the predicted class.
Producer's accuracy %	The count of correctly predicted cases divided by all cases in the actual class.
Kappa	A measure of agreement between the actual and predicted class, ranging from 0 to 1. Kappa may also be weighted to indicate the relative severity of misclassifying a specific class.

There is usually a trade-off between the number of vegetation classes and the degree of misclassification. To improve the accuracy of classification, images are usually obtained when the maximum spectral separation of vegetation classes can be obtained (Daniel *et al* 1998). With any given image, the misclassification rate can be minimized by grouping classes that have similar spectral signatures, forming macro-classes. Although, good classification results are guaranteed, this results in a map that may have a weakened thematic relevance because it may not represent the initial diversity the researcher desires. The ability to identify contrasting vegetation types is usually limited by the sensor's spectral resolution. McKenzie (1999) identified 13 vegetation classes that were associated with the distribution of tuberculosis in possums, but was limited to only classifying eight broader classes due to the spectral resolution of the SPOT3 multispectral image. McKenzie (1999) was interested in discriminating between the different vegetation classes as much as possible since previous research indicated an association between tuberculous possum distribution and different vegetation classes. Boone *et al* (2000) identified 36 vegetation types from a previous image, but were forced to simplify this to eight general vegetation types due to low frequency counts causing problems within the statistical analysis.

When assessing change in land cover data over time, it is necessary to know the misclassification rate of the base vegetation data to understand the practical significance of an observed landscape difference. A small landscape change with large misclassification noise will be meaningless.

2.1.3. Application of Vegetation Classification

In addition to using vegetation classes as variables, researchers have explored mapping of vegetation boundaries where contact between vector and host is maximised (called ecotones), because the habitats are suited to both, as is the case with Lyme disease and the movement of human habitats into forest edges (Beck *et al* 2000).

Vegetation class data has also been used to derive a measure of heterogeneity. Daniel (1998) concluded that vegetation classes associated with heterogeneity of inner structure appeared to offer the best predictive model for tick-borne encephalitis. Heterogeneity of vegetation was also important in McKenzie's (1999) research as a predictor of possum Tb risk.

McKenzie also combined vegetation classes with differing classes representing slope or topography of land, as this combination had been shown to support different possum densities and risk of Tb.

2.2. Habitat Patterns

Although assessment of habitat patterns includes measuring amounts of different vegetation classes within a landscape (mentioned in the previous section), the emphasis in this section is on variables derived from remotely sensed images that describe the shape and spatial interaction of different landscape classes and how this can be related to population density and disease.

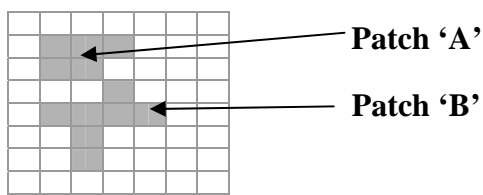
Ecologists have only recently started to explore the relationship between landscape features and population dynamics (Ritchie 1997), especially the importance of heterogeneity and patchiness of the habitat (Pickett and Rogers 1997). Species respond to different landscape patterns in different ways. For example, home ranges of possums are related to heterogeneity of the habitat, with possums located within farmland having a much larger home range than possums located in heterogeneous bush (Effor 2000). The distribution of white-tailed deer is strongly influenced by composition and spatial pattern of resource patches at the landscape level (Roseberry and Woolf 1998). The presence, density, and diversity of birds wintering in fragmented woodlots in Ohio was found to be strongly associated with woodlot area and negatively associated with woodlot isolation from other woodlots (Doherty and Grubb 2000). Therefore measuring habitat patterns can assist in estimation of population dynamics and density, which in turn may improve understanding of the distribution of disease in space and time.

Habitat analysis is a method of describing the spatial structure of the landscape. Habitat patterns can be grouped into three major components; heterogeneity, complexity and scale (McCoy and Bell 1990). Heterogeneity, which is considered the central causal factor in ecological systems (Pickett and Cadenasso 1995), refers to the relative abundance of different patches within a landscape whereas complexity refers to the absolute abundance of each patch. Scale is the size or area at which heterogeneity and complexity are measured. The landscape is the area under study e.g. a farm or geographical region, and a patch is defined as an accumulation of contiguous pixels or a polygon with the same vegetation class. Patches are the basic building block of a landscape pattern.

An image must be 'dissolved' before any habitat variables can be derived. With vector images, this involves removing the boundaries between adjacent patches that have the same vegetation class. With raster images, there are two methods available to 'dissolve' or clump together patches of similar vegetation types, referred to as the four neighbour (4N) and eight neighbour (8N) method.

The 4N method only assigns pixels to the same patch if they touch on any four sides of the pixel. The 8N rule will assign pixels to the same patch if they touch on any four sides and on any diagonal. For example, Figure 1 shows a raster image with potentially two patches of the same vegetation class which are touching on the diagonal. Dissolving the image using the 8N method would result in the two patches being combined into one patch. In Contrast, using the 4N method would result in two distinct patches.

Figure 1 Dissolving patches in raster images - 4N versus 8N rule. The grid shows two patches with the same vegetation class.



2.2.1. Landscape Indices

Landscape indices (or metrics) are used to quantitatively represent habitat patterns. Landscape indices began to be explored in the late 1980's with initially three indices – dominance, contagion and shape (Haines-Young and Chopping 1996). There are now many different indices to choose from but there is still little information available to aid in the selection of appropriate indices for a particular objective, due to a lack of understanding between landscape pattern and process (Haines-Young and Chopping 1996). Riitters *et al* (1995) demonstrated the use of factor analysis in deriving an efficient set of landscape variables, but whether or not landscape variables are biologically useful is dependant on the goals of analysis. For ease of calculation and interpretation, O'Neill *et al* (1999) recommend the following indices, which are relatively independent of each other: average patch perimeter to area ratio; contagion; relative patch area; fractal dimension; and total number of different land cover types. The common indices are briefly discussed below and examples are given of indices that measure patch density, edge, shape, diversity and core area. Examples of area indices are not given, as these are self-explanatory.

2.2.1.1. Contagion

Contagion is the probability that two randomly chosen adjacent pixels belong to the same landscape class and therefore is a measure of the degree of aggregation of the landscape (Hargis *et al* 1997). Values of contagion range between 0 and 100%. Contagion is inversely proportional to edge density

and therefore it is redundant to use both to describe landscape structure. The choice of which index to use depends on the type of research; for example, habitat ecotones are best described by edge density.

2.2.1.2. Mean Nearest-Neighbour Distance

Mean nearest-neighbour distance defines the average inter-patch distance between a patch and its nearest neighbour. This variable is most useful when the proportion of patch is less than 20% of the landscape area. Reporting the standard deviation of this variable gives information about patch dispersion. A small standard deviation implies a fairly uniform spatial distribution of patches, whereas a large value indicates clumping patterns (Hargis *et al* 1997).

2.2.1.3. Mean Patch Fractal Dimension

Fractal dimension characterises the complexity of the patch shapes within the landscape i.e. the jaggedness of curves or the roughness of surfaces. Mean patch fractal dimension approaches one for simple shapes and two for more complex shapes. This index can be weighted (Weighted Mean Patch Fractal Dimension) to control for patch size as large patches are naturally more complex than smaller patches.

2.2.1.4. Edge Density

Edge density is the total length of patch edge per unit area. It can be measured at either the patch or landscape level. Edge density may be useful when patches represent only a small proportion of the landscape. Low ranges of values are produced by large patch landscapes.

2.2.1.5. Mean Proximity Distance

The proximity or isolation indices use a specified search radius to measure the isolation of a patch within a complex of patches. The mean proximity distance is the average proximity index for all patches within the landscape. Highest values occur when patches of the same class are large and close together.

2.2.1.6. Core Area Density

The core area of a patch is the interior area of a patch that is defined by a core area buffer distance. Core area density is the number of disjoint core areas per hectare of total landscape area. The core area is sensitive to the shape of patches. Core area indices may be of most use where species are affected by edge effects, that is, the ecology of a species behaves differently along patch edges than interior patches (McGarigal and Marks 1995).

2.2.2. **Scale**

Scale refers to the resolution at which patterns are measured, perceived, or represented (Milne 1997). Scale has two meanings. Firstly, it refers to the spatial resolution of the map and secondly to the size of the landscape being measured (extent). Most habitat variables are sensitive to scale, that is, a landscape of 100 by 100 metres will have a different quantitative habitat description than a 200 by 200 metre landscape. Therefore the way we interpret heterogeneity and homogeneity of the landscape is dependant on the scale at which it is measured (Pickett and Cadenasso 1995).

In terms of spatial resolution, coarsely resolved habitat maps combine 'suitable' with 'unsuitable' locations, leading to errors of omission and commission (Milne 1997). With respect to landscape extent, the problem with scale can easily be illustrated by measuring the length of a coastline. In order to answer the question of how many nesting sites of bald eagles can fit along the coast of Admiralty Island, the length of the coastline needed to be estimated (Milne, 1997). When applying callipers to an aerial photograph, coastline length values of 380 and 580km were obtained with callipers set at 10km or 3km respectively to step out the coastline length. Therefore, at which scale should landscape complexity be measured? Measurements of landscape can be made at a scale that is of ecological significance to the species of interest, such as species home range length or area. In the above example, Milne (1997) simply used a scale that represented the birds average inter-nest distance (0.782km). Measurements made at multiple scales may be required to characterise the ways in which species respond to spatial complexity (Milne 1997).

2.2.3. **Application of Landscape Indices**

One of the advantages of landscape indices is that they are easy to calculate but they may lack explicit relationships to ecological outcomes of interest (Vos *et al* 2001). Although many examples can be given of the application of habitat pattern analysis to landscape ecology (Frohn 1998), its application in epidemiology is new. Most available literature has used landscape indices to describe

landscape fragmentation and degradation due to human intervention, e.g. deforestation. McKenzie (1999) used landscape indices to describe landscape differences in relation to the risk of farms being positive or negative for wildlife associated tuberculosis (Tb). Landscape indices used included areal characteristics, shape, edge characteristics, core area metrics and richness. Combining vegetation classes with a Data Elevation Model (DEM), McKenzie (1999) further categorised vegetation classes into vegetation-slope classes. DEM's have been used to describe the variability in terrain. Proximity variables have also been useful, such as distance to forest (McKenzie 1999), distance to water, and distance to complimentary habitat components (Osbourne *et al* 2001).

2.3. Vegetation Indices

There are many indices that can be derived from remotely sensed images. Indices other than Vegetation Indices (VI) include Land Surface Temperatures (LST), Atmospheric Moisture Indices, and Rainfall indices. Vegetation indices (also called Spectral Vegetation Indices (SVI)) measure the amount of photosynthetic activity of the landscape. Healthy photosynthesising plants look darker in visible wavelengths and brighter in the near infrared regions than unhealthy plants. Also, dense vegetation coverage absorbs more red and reflects more near infrared radiation than less dense vegetation (Hay 2000). Soils exhibit different spectral properties than vegetation and therefore VIs are designed to maximise the difference in contrast between soils and vegetation to identify the presence of vegetation in RS images. VIs have been applied to predict Rift Valley fever epidemics in Africa, tsetse fly ecology and distribution in West Africa, and tick-borne diseases (Hay *et al* 1997). Vegetation indices at a local scale are usually related to environmental conditions whereas at the regional scale it is often related to broad land cover types (Curran *et al* 2000).

VIs are derived mostly from low spatial / high temporal resolution satellite images (Thomson and Conner 2000). Several different VIs exist: Ratio Vegetation Index (RVI), Normalised Difference Vegetation Index (NDVI), Soil-Adjusted Vegetation Index (SAVI), and Perpendicular Vegetation Index (PVI) (Jackson and Huete 1991). Multi-temporal NDVI data, derived from the NOAA-NVHRR satellite sensor, is the most common method for assessing land-cover change over large areas (Rogers; Hay, and Packer 1997) and other methods are less commonly used. VIs can be used in isolation or in combination with each other for example, Rogers *et al* (1997) used a combination

of different vegetation indexes (NDVI, SAVI, RVI), in conjunction with a Data Elevation Model (DEM) and temperature, to map vegetation cover over Nigeria.

2.3.1. Description of Vegetation Indices

VIs can be measured by ratioing, differencing, ratioing differences and sums, and by forming linear combinations of spectral band data. Developments and improvements in VIs are an ongoing process (Jackson and Huete 1991). The common VIs are briefly described below.

2.3.1.1. Ratio Vegetation Index (RVI)

RVI takes the ratio of the near-infrared and red radiances ($RVI=NIR/R$). This is the most straightforward vegetation index to derive. RVI can be sensitive to vegetation changes during peak growth but is not very sensitive to vegetation differences when the vegetation cover is sparse (Jackson and Huete 1991).

2.3.1.2. Normalised Difference Vegetation Index (NDVI)

NDVI is the most used in epidemiological applications. It is calculated as the difference in the near-infrared (NIR) and red band divided by their sum and is a measure of photosynthetic activity (Baylis *et al* 1998). NDVI is more sensitive to sparse vegetation than RVI, but is less sensitive at high vegetation cover.

2.3.1.3. PVI

There is generally a linear relationship between Red and NIR reflectance's from bare soils and this relationship is referred to the soil line. The PVI is the perpendicular difference between the measured reflectance and the soil line on a NIR-Red (x-axis, y-axis) reflectance graph. Thus, PVI measures the changes in reflectance of the bare soil due to vegetation and represents the vegetation cover independent of the effects of soil.

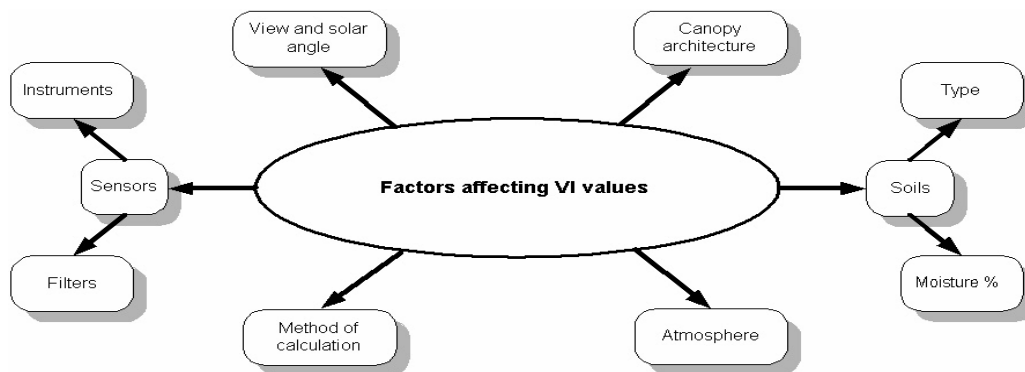
2.3.1.4. Soil-Adjusted Vegetation Index (SAVI)

The structure of the SAVI formula is similar to NDVI, except it has a constant incorporated into the formula to control for the soil background effects. SAVI is an enhancement to PVI.

2.3.2. Factors Affecting Vegetation Indices

There are many factors that affect derivation of VIs, including how they are calculated, see Figure 2. VIs can be calculated from output voltages, reflectance, radiances (ground based measures) or digital numbers (remote sensing). Calculation from these different sources will yield different results. All instruments have unique detectors and filters that measure light within a particular wavelength making measurements of VIs unique to each sensor. As mentioned above, soils will have an influence on VIs until the vegetation completely covers the ground. Wetting of previously dry soils causes soil colour to become darker and soil type also affects soil colour. Red and yellow coloured soils have a significant influence in arid regions and early stages of vegetation growth. Different views and solar angle (position of satellite relative to ground and sun) affect VIs since light reflected from vegetation (non-lambertian surface) is dependent on angle. A sensor that is directly over a vegetation cover (nadir-pointing sensor) can receive more reflected radiation from a canopy that has its leaves in a horizontal plane compared with a vertical position. In general, VIs decrease in value with increasing atmospheric turbidity (Jackson and Huete 1991).

Figure 2 Diagram showing factors affecting the interpretation of vegetation indices.



2.3.3. Application of Vegetation Indices

Researchers have used VIs to derive vegetation classes over a large area, measure changes in vegetation density, or correlate with important environmental variables that affect vector borne diseases.

NDVI values have been correlated with important environmental variables that affect vector borne diseases. Baylis *et al* (1998) compared climatic variables from 68 sites in Algeria to minimum

NDVI of the same sites from the FAO-ARTEMIS CD-ROM. They found relationships between minimum NDVI and total annual rainfall, number of rainy days, minimum humidity, and total annual hours of sunshine. They also showed that average annual rainfall and minimum humidity were significant predictors of minimum NDVI, and average annual rainfall explained 70% of minimum NDVI variation. Linear relationships between annual NDVI and ground based measurement of tsetse fly distribution and cases of sleepy sickness have been documented (Rogers 1991).

Estrada-Pena (1999) provide another example of using VIs as a proxy measure of an environmental variable. The author stated that humidity was an important predictor of tick distribution, but it was not easy to measure either relative humidity or saturation deficit from AVHRR satellite imagery without the help of microwave measurements. In this case, NDVI provide a suitable indirect, and cost effective measure of ambient moisture content.

It is often desirable to measure changes in vegetation cover over time and VIs can be used for this purpose, though measurement is adversely affected by cloud cover. To reduce the effects of cloud cover, the maximum NDVI values may be recorded for each 10-day period, called Maximum Value Composites (MVC). In most cases a combination of NDVI-summarised variables are used collectively as a predictive variable. An example of this summarising is shown in Baylis et al (1998). They used NDVI at a spatial resolution of 7.5km where 3 (decadal) images per month were obtained and combined to produce one monthly representation. These images were further combined to make four annual images: the average annual maximum NDVI; average annual mean NDVI; average annual minimum and the average annual range in NDVI. Another method that is sometimes used is calculating the standard deviation of the monthly or decadal NDVIs to represent the uniformity of the vegetation density (Boone *et al* 2000).

Boone *et al* (2000) were able to use vegetation density and variation of density in combination with GIS, to select sampling sites that gave the widest range of environmental conditions possible. For each type of vegetation, they produced a coloured map where each colour represented a unique combination of high or low vegetation density index, standard deviation of vegetation density index, slope, elevation, and distance from stream. Sampling sites were then randomly selected within each stratum so that they were within 0.5 m of a road and at least 1 km from any other sample site.

The use of individual band and other band ratios have also been explored. Kitron *et al* (1997) examined the correlation of individual bands and band ratios and compared this to the common NDVI, derived from the Landsat 5 Thematic Mapper (TM). They found that bands 7 and 3 had the highest correlation of satellite data with ground measures that were mostly associated with fly catch (% closure, thicket extent, distance from thicket). The correlation between these variables and NDVI was not as high as with the individual band values.

2.4. Vegetation Classes Versus Indices

NDVI are easy to derive and less time consuming than classifying vegetation types. The ability to predict habitat suitability and disease occurrence over a very large area using low spatial resolution images and minimising the effects of cloud cover has proven the usefulness of vegetation indices. With respect to high spatial resolution imagery, researchers have more often classified the vegetation type, either by supervised or non-supervised methods, than using NDVI.

Daniel *et al* (1998) listed reasons why they used a vegetation classification system as opposed to NDVI values. Their reasons were: 1) their study area was typical of a mosaic pattern and large ecotones; 2) the association between suitable plant communities (both species and structure) and tick species had been well studied; 3) *Ixodes ricinus* requires well-defined climate and microclimate conditions which vary in accordance with plant cover change which can be expressed in ecotones and therefore requires high spatial resolution imagery. However, they mention that one draw back of using vegetation classes was the field checking requirement for these classes.

2.5. Other Environmental Variables

Most of the literature on remote sensing applications to epidemiology is dominated by human health issues, and any research on animal health is usually in association with zoonotic effects on humans. A major issue in the past two decades has been investigation of the impact of global warming on epidemiology of diseases for example changes in the distribution of vectors and resulting changes in selected major diseases such as malaria.

Diurnal temperature difference maps can be obtained from the AVHRR sensor on board the NOAA satellite, and are an indication of the surface and subsurface moisture content within the soil and plant canopy (Thompson and Malone 1996). Thompson and Malone (1996) were able to test their

hypothesis that diurnal temperature differences should correlate with the distribution of Bancroftian filariasis cases within villages of the southern Nile Delta, since the intermediate snail host is influenced by environmental moisture factors. Despite using low spatial resolution image (10 km²) and poor village coordinates, they found a high correlation between remotely sensed diurnal temperature differences and prevalence of disease cases.

Flasse *et al* (1998) used Meteosat images as a early warning system to farmers for nasal bot fly in Namibia. The nasal bot fly (*Oestrus ovis*) affects animals in many parts of the world by depositing larvae in the nostrils of small animals. Simple models already exist in Namibia that relate *O. ovis* emergence and soil temperature at a depth of 3cm. Furthermore, an experimental correlation has been established between air degree days and soil degree days. A degree day is equivalent to one degree above the minimum threshold temperature (12°C) for 24 hours. The above information was already in use as an early warning system; when the accumulated soil degree-day reached 265, a warning was given to farmers. However, this system uses only three weather stations to collect information on air degree days for approximately 2000 farms. The Meteosat satellite recorded brightness temperatures every 30 minutes at a spatial resolution of 5km images, and these have been used to assimilate land surface temperatures. Flasse *et al* (1998) showed that the Meteosat derived degree days were within 16 % of the weather station degree day recordings, which may only correspond to 3 - 4 days when related to emergence of *O. ovis*. This variation is acceptable since the larvae remain in the sheep's head for about one month, and the timing of application of parasiticidal drugs can vary by one week with little loss of efficacy. The authors also reported that it was unclear whether the weather station or the satellite information was more accurate as there was no systematic validation of the usefulness of cumulative degree-days. In addition, satellite imagery has the ability to discriminate geographical location, and therefore instead of giving one warning to approximately 2000 farmers, warnings could be given at a much smaller scale.

With the help of remote sensing, Zukowski *et al* (1991) were able to refine their one farm study area (760 ha), relating individual farm risk of liver fluke to a soil based model of snail (*Fassaria bulimoides*) habitat, to 12 farms (798 ha). Their refined model showed a significant relationship to chenier soils with the Hackberry-Mermentau complex, something that was not evident in the first study.

3. Conclusion

Remote sensing provides epidemiologists with a method of obtaining environmental data over a vast geographic area in a cost effective manner. The types of environmental data collected may include those factors that affect vector population dynamics, such as temperature, rainfall, humidity and type of vegetation. Measuring the 'greenness' of vegetation using vegetation indices or classifying vegetation and measuring its spatial complexity have been shown to provide useful correlates of disease dynamics and distribution.

The following study examines the use of satellite derived habitat variables as predictors of wildlife associated Tb.

4. Reference List

- Andrefouet S and Claereboudt M. Objective class definitions using correlation of similarities between remotely sensed and environmental data. *International Journal of Remote Sensing*. 2000; 21(9):1925-1930.
- Baylis M, Bouayoune H, Touti J and Hasnaoui H. Use of climatic data and satellite imagery to model the abundance of *Culicoides imicola*, the vector of African horse sickness virus, in Morocco. *Medical and Veterinary Entomology*. 1998; 12:255-266.
- Beck LR, Lobitz BM and Wood BL. Remote sensing and Human Health: New Sensors and New Opportunities. *Emerging Infectious Diseases*. 2000; 6(3):217-227.
- Booman M, Durrheim DN, La Grange K, Martin C, Mabuza A, Zitha A, Mbokazi F, Frazer C and Sharp B. Using a geographical information system to plan a malaria control programme in South Africa. *Bulletin of the World Health Organisation*. 2000; 78(12):1438-1444.
- Boone JD, McGwire KC, Otteson EW, DeBaca RS, Kuhn EA, Villard P, Brussard PF and St Jeor SC. Remote sensing and geographic information systems: charting Sin Nombre Virus infections in Deer Mice. *Emerging Infectious Diseases*. 2000; 6(3):248-258.

- Cracknell AP and Hayes LWB. Introduction to Remote Sensing. New York: Taylor & Francis; 1991.
- Curran PJ, Atkinson PM, Foody GM and Milton EJ. Linking remote sensing, land cover and disease. *Advances in Parasitology*. 2000. 2000; 47:37-80.
- Daniel M, Kolar J, Zeman P, Pavelka K and Sadlo J. Predictive map of *Ixodes ricinus* high-incidence habitats and a tick-borne encephalitis risk assessment using satellite data. *Experimental & Applied Acarology*. 1998; 22(7):417-433.
- Doherty PF and Grubb TC. Habitat and landscape correlates of presence, density, and species richness of birds wintering in forest fragments in Ohio. *The Wilson Bulletin*. 2000; 112(3):388-394.
- Effor M. Possum Density, Population Structure, and Dynamics. Montague TL, Editor. *The Brushtail Possum*. Lincoln, New Zealand: Manaaki Whenua Press; 2000; pp. 47-61.
- Estrada-Pena A. Geostatistics and remote sensing using NOAA-AVHRR satellite imagery as predictive tools in tick distribution and habitat suitability estimations for *Boophilus microplus* (Acari: Ixodidae) in South America. *Veterinary Parasitology*. 1999; 81:73-82.
- Flasse S, Walker C, Biggs H, Stephenson P and Hutchinson P. Using remote sensing to predict outbreaks of *Oestrus ovis* in Namibia. *Preventive Veterinary Medicine*. 1998; 33:31-38.
- Frohn RC. Remote sensing for landscape ecology: New metric indicators for monitoring, modelling and assessment of ecosystems. Boca Raton: Lewis Publishers; 1998.
- Haines-Young R and Chopping M. Quantifying landscape structure: a review of landscape indices and their application to forested landscapes. *Progress in Physical Geography*. 1996; 20(4):418-445.
- Hargis H, Bissonette J and David J. Understanding measures of landscape pattern. Bissonette JA, Editor. *Wildlife and Landscape Ecology. Effects of pattern and scale*. New York: Springer-Verlag; 1997; pp. 231-261.

- Harrison BA and Jupp DLB. Introduction to remotely sensed data. Australia: CSIRO Publications; 1990.
- Hay SI. An overview of remote sensing and geodesy for epidemiology and public health application. *Advances in Parasitology*. 2000. 2000; 47: 1-35.
- Hay SI, Packer MJ and Rogers DJ. The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. *International Journal of Remote Sensing*. 1997; 18(14):2899-2930
- Jackson RD and Huete AR. Interpreting vegetation indices. *Preventive Veterinary Medicine*. 1991; 11:185-200.
- Kitron U, Otieno LH, Hungerford LL, Odulaja A, Brigham WU, Okello OO, Joselyn M, Mohamed-Ahmed MM and Cook E. Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. *Journal of Animal Ecology*. 1996; 65:371-380.
- McCoy ED and Bell SS. Habitat structure: The evolution and diversification of a complex topic. In: Turner MG and Gardner RH, Editors. *Quantitative Methods in Landscape Ecology*. New York: Springer-Verlag ; 1990.
- McGarigal K and Marks B. FRAGSTATS: Spatial analysis program for quantifying landscape structure. U.S. Department of Agriculture Forest Service General Technical Report PNW-GTR-351; 1995.
- McKenzie JS. The use of habitat analysis in the control of wildlife tuberculosis in New Zealand. Palmerston North, New Zealand: PhD Thesis. Massey University; 1999.
- Mills JN and Childs JE. Ecological studies of rodent reservoirs: Their relevance for human health. *Emerging Infectious Diseases*. 1998; 4(4):529-538.
- Milne BT. Applications of Fractal Geometry in Wildlife Biology. Bissonette JA, Editor. *Wildlife*

and Landscape Ecology. Effects of pattern and scale. New York: Springer-Verlag; 1997; pp. 32-69.

O'Neill R, Riitters K, Wickham J and Jones K. Landscape pattern metrics and regional assessment. *Ecosystem Health*. 1999; 5(4):225-233.

Osbourne PE, Alonso JC and Bryant RG. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology*. 2001; 38:458-471.

Pickett STA and Cadenasso ML. Landscape ecology: spatial heterogeneity in ecological systems. *Science*. 1995; 269(5222):331-334.

Pickett STA and Rogers KH. Patch dynamics: the transformation of landscape structure and function. Bissonette JA, Editor. *Wildlife and Landscape Ecology. Effects of pattern and scale*. New York: Springer-Verlag; 1997; pp. 101-127.

Riitters KH, O'Neill RV, Hunsaker CT, Wickham JD, Yankee DH, Timmins SP, Jones KB and Jackson BL. A factor analysis of landscape pattern and structure metrics. *Landscape Ecology*. 1995; 10:23-39.

Ritchie ME. Populations in a Landscape Context: Sources, Sinks, and Metapopulations. Bissonette JA, Editor. *Wildlife and Landscape Ecology. Effects of pattern and scale*. New York: Springer-Verlag; 1997; pp. 160-184.

Rogers DJ. Satellite imagery, tsetse and trypanosomiasis in Africa. *Preventive Veterinary Medicine*. 1991; 11(3-4):201-220.

Rogers DJ, Hay SI and Packer MJ. Mapping land-cover over large areas using multispectral data derived from the NOAA-AVHRR: a case study of Nigeria. *International Journal of Remote Sensing*. 1997; 18(15):3297-3303.

Roseberry JL and Woolf A. Habitat-population density relationships for white-tailed deer in Illinois. *Wildlife Society Bulletin*. 1998; 26(2):252-258.

- Thompson DF and Malone JB. Bancroftian filariasis distribution and diurnal temperature differences in the Southern Nile Delta. *Emerging Infectious Diseases*. 1996; 2(3):234-345.
- Thomson MC and Conner J. Environmental information systems for the control of arthropod vectors of disease. *Medical and Veterinary Entomology*. 2000; 14:227-244.
- Vos CC, Verboon J, Opdam PFM and Ter Braak CJF. Toward Ecologically Scaled Landscape Indices. *The American Naturalist*. 2001; 183(1):24-41.
- Zukowski SH, Hill JM, Jones FW and Malone JB. Development and validation of a soil-based geographic information system model of habitat of *Fossaria bulimoides*, a snail intermediate host of *Fasciola hepatica*. *Preventive Veterinary Medicine*. 1991; 11:221-227.

CHAPTER 2: USE OF REMOTE SENSING TO CREATE A PREDICTIVE MODEL OF BOVINE TUBERCULOSIS

1. Introduction

The usefulness of remote sensing (RS) in epidemiology is related to the correlation between mapped environmental variables and disease dynamics and distributions (Boone *et al* 2000). Vegetation classification, habitat patterns, and vegetation indices have been the most commonly derived remotely sensed variables, which have been applied to the study of animal and human diseases. Vegetation classification involves creating a digital image allows identification of specific vegetation types. These vegetation types can either be homogeneous in composition (e.g. pine forest or grassland) or heterogeneous e.g. a mixture of different classification types. The distribution of disease reservoir and host may relate to the occurrence and distribution of specific vegetation types rather than broad vegetation cover (Beck *et al* 2000). For example the distribution of certain plants may affect the microclimate for ticks and the living conditions of their hosts influencing favourable food resources and shelters (Daniel *et al* 1998; Mills and Childs 1998).

In contrast to the field of landscape ecology (Frohn 1998), the application of habitat pattern analysis to the field of veterinary epidemiology is relatively new. McKenzie (1999) used landscape indices to explain differences in the risk of farms being positive to wildlife associated tuberculosis (Tb). Bovine tuberculosis, caused by *Mycobacterium bovis*, is an endemic disease within the possum population of New Zealand. The brushtail possum (*Trichosurus vulpecula*) is an important wildlife vector for bovine Tb in New Zealand, therefore a significant investment is directed towards the control of wild possum populations. By 2013, the National Pest Management Strategy for Bovine Tuberculosis aims to reduce the annual period prevalence in cattle and deer herds to fewer than 0.2%. The prevalence of infected cattle herds at 30 June 2002 was 0.5% (Animal Health Board).

Currently control of Tb in wildlife relies primarily on culling of possums. The distribution of possums in New Zealand is habitat dependent (Effor 2000) and methods that allow identification of habitats associated with high density possum populations, may allow more efficient allocation of resources for possum control.

This paper aimed to determine if satellite derived habitat variables could be used as indirect indicators of wildlife associated Tb by associating them in a statistical model with known farm level Tb status.

2. Materials and Methods

2.1. Vegetation Image

Vegetation data was derived from a SPOT4 multispectral satellite image. The image was captured in 1999 and covered a 6949 km² area of the Wairarapa. Each pixel represented a 20 x 20 metre area of land. The image was classified into 13 different vegetation classes using unsupervised classification performed by staff at LandCare: Podocarp/Broadleaf; Broadleaf scrub/Pure Broadleaf; Broadleaf scrub / Manuka; Pine / Mixed; Willows; Manuka / Gorse; Pine; Manuka; Gorse; Crop; Pasture; Water; Bare and Undefined. The last four classes were excluded from the analysis as they have been associated with very low possum densities (Effor 2000).

2.2. Calculation of Habitat Variables

Patch Analyst® (Elkie; Rempel and Carr 2003) was used to calculate habitat variables for each vegetation class within each farm. Farm boundaries were overlaid on the vegetation image.

Patch Analyst was used to generate habitat variables for each landscape, of which 12 were considered to be of practical importance and were retained for analysis (Table 3). Variables that were included were those that measured area, variation in size, distance between patches, and whether patches tended to be clumped together or uniformly dispersed throughout a farm. Habitat variables not considered were mostly those that represented shape of habitat patches and were not considered to be biologically important. Variables that were presented as both a percentage of landscape (farm area) and as area were retained in percentage form for analysis. This was to help reduce the effect of differing landscape size across the study.

Table 3. List of habitat variables calculated from Patch Analyst®.

Code	Name	Description
FarmArea	Farm Area	Total land area of a farm in hectares.
Farm_IJI	Interspersion Juxtaposition Index	A measure of the interspersion of each patch in the landscape.
CACoV	Core Area Coefficient of Variation	Measures the variation in the core area's over a landscape
CLAND	Core Area of Land	The area covered by pixels within a patch that are at least 40 metres from any part of the patch's perimeter.
CAD	Core Area Density	Number of disjunct core areas per hectare.
ED	Edge Density	Amount of edge relative to the total area of the vegetation class within the farm (metres/hectare).
LPI	Largest Patch Index	Largest patch of a specified class within the farm boundary.
MCA	Mean Core Area	Average core area per patch
MNN	Mean Nearest Neighbour	Mean of the shortest distances to a similar patch for each patch.
MPI	Mean Proximity Index	Measure of the degree of isolation and fragmentation
MPS	Mean Patch Size	Mean size of patches in hectares.
Percent of Land	Land Percent	Percentage of total farm area occupied by the vegetation class.

2.3. Selection of Farms

The Wairarapa area was chosen as Tb is known to be endemic within the wild possum population. A vector map of farm boundaries for the study area was obtained from Agribase. Only beef breeding farms were selected because this farming system is less likely to graze animals off the farm and therefore the surrounding habitat is most likely to represent the Tb conditions on farm. Farms were eligible for inclusion in the study if they consisted of a single block in Agribase and had a single herd identity recorded in the National Livestock Database (NLDB). An approximate total Livestock Units (LSU) for each farm was calculated from Agribase records and then compared to the total area of the farm. Farms with an unusually high stocking rate (> 15 LSU per hectare) were excluded.

2.4. Dependent Variable

The dependent variable was calculated as the proportion of years each farm was positive to Tb. Historical Tb data from the National Livestock Database (NLDB), administered by AgriQuality was obtained for all farms in the Wairarapa covering the period from 1979 to 1999. The NLDB stores information about Tb testing on farms as required by the national Tb control programme. The Tb data also included when and where possum control first started. Farms were eligible for selection if they had at least 10 years of Tb data before any Tb control programme was initiated in the area. A 10 year interval was felt to be necessary to have an appreciation of the Tb status of a farm over time as Tb within possums fluctuates over time and having a long interval ensured Tb hotspots were likely to be represented.

A farm was considered positive to Tb in any one year when at least one animal was confirmed positive to the tuberculin test. The proportion of years positive to Tb was calculated as the sum of years positive to Tb divided by the total number of years which a farm was tested before the initial control.

2.5. Statistical Analysis

Univariate and bivariate analyses were performed on all habitat variables. Density trace graphs were used to assess the distribution of variables. Variables with highly skewed distributions were categorised to prevent instability in the multivariable model. Variables with very little variation were excluded from further analysis.

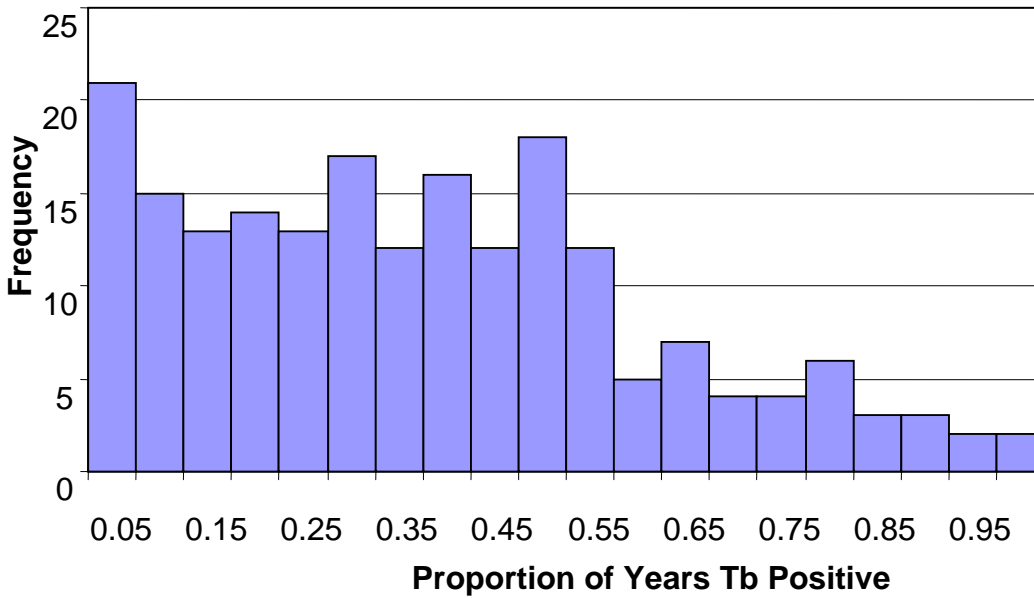
Bivariate screening analyses were done to test for associations between each habitat variable within each vegetation class, and the outcome variable. Most of the habitat variables produced from Patch Analyst were highly significant in the bivariate analysis and a bivariate correlation matrix demonstrated multicollinearity. For each vegetation class, Principle Components Analysis (PCA) was then used to select a subset of orthogonal habitat variables for consideration in a subsequent multivariable model. The variable with the highest component score was selected for each principle component that had an eigenvalue greater than one (Tabachnick and Fidell 1996). A separate multivariable model was created for each of the nine vegetation classes. Logistic regression was used to analyse the relationship between independent variables and proportion of years Tb positive within each vegetation class using the GENMOD procedure in SAS (SAS Institute 1998) using a manual backward selection process.

Continuous variables present in the final models were examined for linearity in the logit as described by Hosmer and Lemeshow (2000) and variables were categorised if the logit did not present a linear trend. An attempt was then made to combine information for different vegetation classes in one logistic regression model. Variables from the separate vegetation class models were considered as blocks in the model with farm_IJI and farm area forced in. This model was built using a forward stepwise process (inclusion threshold = 0.05; exclusion threshold=0.10), with selection or exclusion of the separate class models done at the block level. Model checking was performed as described by SAS Institute Inc (1995) and Allison (1999).

3. Results

The study area comprised 434605 ha of predominantly farming land in the Wairarapa, bordered by the Tararua mountains to the south west and the coastline to the east. A total of 2164 farms were identified in the study area. Of these, 199 were included in the study because they met eligibility requirements. Median farm size was 503 ha (range: 10 ha to 2609 ha) with the total area of all 199 farms being 112227 ha. Each farm contributed at least 10 years of farm level Tb data with the range being 10 to 20 farm-years. 183 farms (92%) had at least one year classed as Tb positive. Figure 3 shows the distribution of the dependant variable, proportion of years Tb positive, which fitted a binomial distribution.

Figure 3 Histogram showing variation of the dependant variable: proportion of years positive to Tb



The nine different habitat classes accounted for 4.4% of the total land cover over the 199 study farms. The proportion of farmland covered by habitat varied from 0.4% to 94.7% with a median of 12.1%, see Table 4. All farms had at least some habitat coverage other than pasture but not all habitat classes were necessarily present on all farms. Seventy two percent of the farms contained all nine habitat classes, a further 18% of farms contained six to eight habitat classes and the rest (10%) contained 1 to 5 habitat classes. The maximum percentage of habitat cover for pine, pine /mixed and the maximum for all habitat classes combined was due to one farm. This farm was 284ha and had 15 ha of pasture. Manuka, willows and gorse habitat classes accounted for 65% of the total habitat cover and 2.83% of the land cover.

Table 4 Descriptive statistics showing the percentage of land covered by the different habitat classes for each farm over the 199 study farms. Bottom row shows variation in total habitat across farms.

Habitat Type	Percentage of Land Cover			
	Median	Mean	Minimum	Maximum
Broadleaf Scrub / Pure	0.13	0.34	0.00	3.37
Pine	0.25	1.70	0.00	59.77
Broadleaf Scrub / Manuka	0.26	0.54	0.01	4.57
Gorse	0.65	1.05	0.01	11.79
Pine / Mixed	0.67	1.59	0.00	16.77
Manuka / Gorse	0.69	1.55	0.01	11.73
Podocarp / Broadleaf	1.27	2.48	0.02	27.34
Willows	1.73	2.42	0.07	10.75
Manuka	4.43	5.72	0.05	30.09
All Habitats Combined	12.14	16.41	0.43	94.69

Each of the nine multivariable models developed at the vegetation class level produced significant risk factors. When these interim models were considered for inclusion as blocks into a final multivariable model, only six class level blocks were retained. The results of the final multivariable model incorporating two farm level variables and six blocks of vegetation class level variables, are presented in Table 5.

Table 5 Combined multivariable logistic regression results. Significant variables in each of the class models were entered into the model as a block. Only classes that significantly improved the model fit were retained.

Parameter	Coefficient	SE ¹	DF ²	OR ³	95% CI ⁴		P value
					Lower	Upper	
Intercept	-3.129	0.423	1	0.04	0.02	0.10	<.0001
<i>Farm Variables</i>							
Farm Area (50 Ha Units)	0.0390	0.003	1	1.04	1.01	1.02	<.0001
Farm_IJI (10% Units)	0.1075	0.045	1	1.11	1.02	1.22	0.0173
<i>Podocarp / Broadleaf</i>							
Percent of Land	0.0128	0.020	1	1.01	0.97	1.05	0.5231
MPS (0 – 0.15 Ha)			0	1.00			
MPS (0.16 – 0.23 Ha)	0.3805	0.108	1	1.46	1.18	1.81	0.0004
MPS (> 0.24 Ha)	0.3963	0.136	1	1.49	1.14	1.94	0.0036
ED	0.0001	0.001	1	1.00	1.00	1.00	0.9259
<i>Broadleaf Scrub / Pure</i>							
Percent of Land	0.044	0.121	1	1.04	0.82	1.33	0.7164
MNN (0 – 95 m)			0	1.00			
MNN (95 – 226 m)	0.3871	0.126	1	1.47	1.15	1.89	0.0021
MNN (> 266 m)	0.327	0.136	1	1.39	1.06	1.81	0.0165
<i>Manuka / Gorse</i>							
Percent of Land	0.0501	0.033	1	1.05	0.99	1.12	0.1289
CAD (Absent)			0	1.00			
CAD (Present)	0.6366	0.153	1	1.89	1.40	2.55	<.0001
<i>Pine</i>							
Percent of Land	-0.0094	0.019	1	0.99	0.95	1.03	0.6089
MPS	0.1383	0.129	1	1.15	0.89	1.48	0.2821
MNN (0 – 151m)			0	1.00			
MNN (151 – 292m)	0.5048	0.121	1	1.66	1.31	2.10	<.0001
MNN (> 292 m)	0.3844	0.115	1	1.47	1.17	1.84	0.0008
<i>Manuka</i>							
Percent of Land	-0.0047	0.021	1	1.00	0.96	1.04	0.8209
MPS (1 Ha)	0.865	0.394	1	2.38	1.09	5.13	0.0281

¹Standard Error

²Degrees of freedom

³Odds ratio

⁴95% upper and lower confidence intervals

An odds ratio (OR) greater than one for a given variable indicates an increase in the odds of a TB-positive year as the variable increases in magnitude. An OR less than one variable indicates a decrease in the odds of a TB-positive year as the variable increases in magnitude.

Both farm level variables, Farm_IJI and Farm Area, were statistically significant. Larger farms had a small increased risk of Tb infection and farms with a higher level of interspersion of habitat patches was also associated with higher Tb risk. The percentage of farm area occupied by each of the habitat classes were not significant in the final model, although it was a significant risk factor in the interim class model. However, the presence of Manuka /Gorse patches that contained an area within the patch which was at least 40 metres from any part of the patches perimeter was associated with a 1.9 fold increase in the risk of a Tb-positive year. A 2.4 fold increase in the risk of a Tb-positive year was associated with a one hectare increase in the mean size of Manuka patches, although the overall proportion of Manuka patches present on a farm was not associated with Tb risk.

4. Discussion

The purpose of this study was to indirectly correlate habitat structure on individual farms to possible Tb-positive possum populations. Habitat structure at the farm level was quantitatively measured by overlaying a satellite image on a digital farm boundary map to create individual farm habitat maps from which habitat indices were calculated.

It was assumed that herd infection was mostly due to transmission of Tb from wildlife possum populations to cattle and therefore the presence of Tb-positive farms was associated with the presence of Tb positive possum populations or hotspots. Hotspots may not necessarily be active every year, so to ensure that any hotspots within farms are represented, at least 10 years of Tb data was used. Using many years of Tb data presents another problem since the satellite image reflected the vegetation cover in one point in time (1999), whereas historic farm Tb data was drawn from a 20-year period between 1979 and 1999. The correlation of habitat variables to farm Tb status assumed that the vegetation in the last 20 years was constant.

The variables included in the final model were those that were significant in the individual class models. This model is therefore not parsimonious mainly because it contained non-significant variables. The major problem in this analysis was identifying a set of variables that were orthogonal and best represented each of the vegetation classes. It was already known before the analysis that there would likely be a problem with multicollinearity that would not necessarily be identified by simple bivariate correlations. PCA was a suitable approach to identifying a set of orthogonal

variables for each habitat class, but did not exclude the risk of multicollinearity in the final model when the different vegetation classes were combined.

Variables that represented the percentage of land for each of the habitat classes were not statistically significant in the overall model, even though they were significant in the interim models. Farms with a higher risk of Tb tended to be larger farms, had larger patches of Podocarp/Broadleaf and Manuka, and the distance between Pine neighbouring Pine patches was greater. Core area measurements were included in the model to measure those patches that were a reasonable size and were not 'long and narrow' like patches following river banks. Farms that contained core areas of Manuka/Gorse were associated with an increased risk of a Tb-positive year.

McKenzie (1999) also combined the results of individual habitat class models into one multivariable model. The interspersion of Manuka/Gorse, number of patches of Pine and number of patches of Manuka/Gorse were positively associated with Tb-positive years, whereas the core area density of Shrubland and the total edge surrounding Podocarp/broadleaf patches were negatively associated with Tb-positive years. The present study did not find any habitat variables that were associated with a decreased risk of Tb-positive years. Although this study was based on the same study area as McKenzie's study, it used a vegetation image with a higher spectral resolution which resulted in more habitat classes being classified, and used a different approach to selecting habitat variables at each of the class levels.

One of the major problems with using landscape metrics to make habitat comparisons between farms is that they are affected by the scale at which they are measured. Farm size was not constant across the study and therefore some bias may be introduced, but this was minimised by excluding those metrics which were most sensitive to scale, namely shape metrics (Saura and Martinuz-Millan 2001). Furthermore, it was decided to use farm area as it offered a more practical solution since a digital map of farm boundaries was already available, and the dependant variable (number of animals Tb positive) was able to be summarised at the farm level. Therefore, habitat variables were produced at a different scale for each farm.

Percentage of farm area was used to represent the amount of vegetation on farm instead of absolute vegetation area. This method has the advantage of adjusting the influence of each vegetation class on the Tb status of each farm such that vegetation patterns on small farms are not over represented. However, it does not take into account that a minimum habitat size may be required to support a possum population.

The outward spread of infection from infected possum populations to previous uninfected populations has not been prevented by current control methods (McKenzie *et al* 2002). The success of relating habitat structure to the density of possum populations would allow possum control to be targeted at the farms which are currently free of Tb infection but have a high risk of becoming infected in the near future.

Formulating a predictive model of Tb using variables derived from vegetation images is difficult as there is little understanding of the role these variables provide in describing possum habitat. These variables are also correlated at a very deep level and even though an effort can be made to screen out bivariate correlations, the correlations between groups of variables becomes harder to identify and deal with. It is extremely difficult to describe the role of each individual variable in the final model and thus a model involving habitat variables should only be used to predict an outcome rather than describe the biology.

5. Acknowledgements

I would like to thank AgriQuality for providing the digital vector map of farm boundaries and Landcare Research for the SPOT4 multispectral satellite image.

6. References

- Allison PD. Logistic Regression Using the SAS System. Theory and Application. Cary, NC, USA: SAS Institute Inc; 1999.
- Animal Health Board. Animal Health Board Annual Report 2001/02.
- Beck LR, Lobitz BM and Wood BL. Remote sensing and Human Health: New Sensors and New Opportunities. *Emerging Infectious Diseases*. 2000; 6(3):217-227.
- Boone JD, McGwire KC, Otteson EW, DeBaca RS, Kuhn EA, Villard P, Brussard PF and St Jeor SC. Remote sensing and geographic information systems: charting Sin Nombre Virus infections in Deer Mice. *Emerging Infectious Diseases*. 2000; 6(3):248-258.
- Daniel M, Kolar J, Zeman P, Pavelka K and Sadlo J. Predictive map of *Ixodes ricinus* high-

- incidence habitats and a tick-borne encephalitis risk assessment using satellite data. *Experimental & Applied Acarology*. 1998; 22(7):417-433.
- Effor M. Possum Density, Population Structure, and Dynamics. Montague TL, Editor. *The Brushtail Possum*. Lincoln, New Zealand: Manaaki Whenua Press; 2000; pp. 47-61.
- Elkie P, Rempel R, and Carr A. *Patch Analyst. Users Manual*. Canada; 2003.
- Frohn RC. *Remote sensing for landscape ecology: New metric indicators for monitoring, modelling and assessment of ecosystems*. Boca Raton: Lewis Publishers; 1998.
- Hosmer DW and Lemeshow S. *Applied Logistic Regression*. USA.: Wiley-Interscience; 2000.
- McKenzie JS. *The use of habitat analysis in the control of wildlife tuberculosis in New Zealand*. Palmerston North, New Zealand: Massey University; 1999.
- McKenzie JS, Morris RS, Pfeiffer DU and Dymond JR. Application of remote sensing to enhance the control of wildlife-associated *Mycobacterium bovis* infection. *Photogrammetric Engineering & Remote Sensing*. 2002; February:153-159.
- Mills JN and Childs JE. Ecological studies of rodent reservoirs: Their relevance for human health. *Emerging Infectious Diseases*. 1998; 4(4):529-538.
- SAS Institute. *SAS/STAT ® User's Guide, Volume 6*. U.S.A.: SAS Publishing; 1998.
- SAS Institute Inc. *Logistic Regression Using the SAS System*. Cary, NC, USA: SAS Institute Inc; 1995.
- Saura S and Martinuz-Millan J. Sensitivity of landscape pattern metrics to map spatial extent. *Photogrammetric Engineering & Remote Sensing*. 2001; 67(9):1027-1036.
- Tabachnick BG and Fidell LS. *Using Multivariate Statistics*. New York: HarperCollins College Publishers ; 1996.

APPENDICES

1. Appendix A: Principle Components Analysis

Appendix A reports the results from the Principle Component Analysis for each vegetation class.

1.1. Podocarp / Broadleaf

Total Variance Explained^a

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.721	47.674	47.674	5.721	47.674	47.674
2	1.888	15.735	63.409	1.888	15.735	63.409
3	1.171	9.759	73.168	1.171	9.759	73.168
4	.748	6.237	79.405			
5	.630	5.251	84.656			
6	.562	4.682	89.339			
7	.326	2.718	92.056			
8	.311	2.588	94.645			
9	.248	2.068	96.712			
10	.183	1.527	98.239			
11	.132	1.102	99.342			
12	7.901E-02	.658	100.000			

Extraction Method: Principal Component Analysis.

a. Only cases for which Class = 2 are used in the analysis phase.

Component Matrix^{a,b}

	Component		
	1	2	3
Landpct	.836	.116	-.068
CACoV	.799	-.269	.182
CAD	.837	.179	.143
CLAND	.785	.234	.188
ED	.288	.703	-.457
IJI	.227	-.640	-.405
LPI	.550	.658	-.203
MCA	.728	-.198	.182
MNN	-.280	.261	.781
MPI	.830	-.111	.130
MPS	.922	2.972E-02	-.003
PSCoV	.697	-.507	-.057

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

b. Only cases for which Class = 2 are used in the analysis phase.

1.2. Broadleaf Scrub / Pure

Total Variance Explained^a

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.323	52.693	52.693	6.323	52.693	52.693
2	1.639	13.657	66.350	1.639	13.657	66.350
3	1.200	10.003	76.353	1.200	10.003	76.353
4	.871	7.257	83.610			
5	.583	4.855	88.465			
6	.539	4.496	92.960			
7	.391	3.258	96.218			
8	.204	1.700	97.918			
9	.157	1.311	99.228			
10	7.959E-02	.663	99.892			
11	1.096E-02	9.132E-02	99.983			
12	2.039E-03	1.699E-02	100.000			

Extraction Method: Principal Component Analysis.

a. Only cases for which Class = 4 are used in the analysis phase.

Component Matrix^{a,b}

	Component		
	1	2	3
Landpct	.622	-.457	.263
CACoV	.873	.278	-.098
CAD	.911	.290	-.083
CLAND	.909	.347	-.064
ED	.491	-.742	6.024E-02
IJI	-.082	.300	.864
LPI	.690	-.398	-.362
MCA	.767	.323	-.018
MNN	-.142	.428	-.321
MPI	.892	.149	-.007
MPS	.833	-.194	3.074E-02
PSCoV	.834	-.040	.354

Extraction Method: Principal Component Analysis.

- a. 3 components extracted.
b. Only cases for which Class = 4 are used in the analysis phase.

1.3. Broadleaf Scrub / Manuka

Total Variance Explained^f

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.947	41.228	41.228	4.947	41.228	41.228
2	1.872	15.597	56.826	1.872	15.597	56.826
3	1.626	13.551	70.377	1.626	13.551	70.377
4	1.088	9.064	79.440	1.088	9.064	79.440
5	.888	7.396	86.836			
6	.583	4.862	91.698			
7	.463	3.862	95.560			
8	.205	1.709	97.269			
9	.128	1.064	98.334			
10	.115	.959	99.292			
11	6.763E-02	.564	99.856			
12	1.732E-02	.144	100.000			

Extraction Method: Principal Component Analysis.

- a. Only cases for which Class = 5 are used in the analysis phase.

Component Matrix^{a,b}

	Component			
	1	2	3	4
Landpct	.611	-.555	2.719E-02	-.076
CACoV	.533	1.130E-02	.204	-.745
CAD	.694	.603	-.018	.206
CLAND	.653	.686	-.026	.100
ED	.604	-.323	-.570	.143
IJI	.290	-.214	.655	.464
LPI	.428	2.290E-02	-.777	.161
MCA	.801	.534	7.377E-02	-.189
MNN	-.301	.321	.334	9.971E-02
MPI	.853	-.306	.198	-.224
MPS	.810	-.235	8.204E-02	6.017E-02
PSCoV	.797	-.208	.247	.338

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

b. Only cases for which Class = 5 are used in the analysis phase.

1.4. Pine / Mixed

Total Variance Explained^f

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.969	49.738	49.738	5.969	49.738	49.738
2	1.483	12.354	62.092	1.483	12.354	62.092
3	1.122	9.352	71.444	1.122	9.352	71.444
4	.896	7.470	78.914			
5	.728	6.069	84.984			
6	.642	5.352	90.336			
7	.397	3.308	93.644			
8	.242	2.018	95.662			
9	.196	1.631	97.293			
10	.146	1.219	98.513			
11	.115	.960	99.473			
12	6.326E-02	.527	100.000			

Extraction Method: Principal Component Analysis.

a. Only cases for which Class = 7 are used in the analysis phase.

Component Matrix^{a,b}

	Component		
	1	2	3
Landpct	.741	-.324	6.464E-02
CACoV	.786	-.001	.137
CAD	.759	.170	.206
CLAND	.642	.688	-.157
ED	.771	-.344	3.616E-02
IJI	.262	-.289	-.696
LPI	.789	6.643E-02	.227
MCA	.485	.706	-.340
MNN	-.233	.298	.530
MPI	.849	-.262	.206
MPS	.908	3.869E-03	7.226E-02
PSCoV	.832	-.114	-.226

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

b. Only cases for which Class = 7 are used in the analysis phase.

1.5. Willows

Total Variance Explained^f

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.374	44.783	44.783	5.374	44.783	44.783
2	2.028	16.900	61.683	2.028	16.900	61.683
3	1.299	10.822	72.505	1.299	10.822	72.505
4	1.005	8.371	80.876	1.005	8.371	80.876
5	.820	6.833	87.709			
6	.457	3.804	91.513			
7	.352	2.930	94.443			
8	.243	2.025	96.468			
9	.181	1.505	97.973			
10	.128	1.070	99.043			
11	8.230E-02	.686	99.729			
12	3.254E-02	.271	100.000			

Extraction Method: Principal Component Analysis.

a. Only cases for which Class = 9 are used in the analysis phase.

Component Matrix^{a,b}

	Component			
	1	2	3	4
Landpct	.708	-.383	.246	.239
CACoV	.764	-.200	6.997E-02	-.461
CAD	.836	1.498E-02	3.909E-02	-.222
CLAND	.623	.570	-.477	.168
ED	.269	.320	.794	.265
IJI	.101	-.485	-.230	.440
LPI	.519	.669	.285	.190
MCA	.686	.499	-.464	.128
MNN	-.304	.556	.103	-.487
MPI	.849	-.302	8.173E-02	-.182
MPS	.929	-.006	4.626E-02	.158
PSCoV	.839	-.285	-.076	-.214

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

b. Only cases for which Class = 9 are used in the analysis phase.

1.6. Manuka / Gorse

Total Variance Explained^a

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.117	59.312	59.312	7.117	59.312	59.312
2	1.319	10.992	70.304	1.319	10.992	70.304
3	.936	7.801	78.104			
4	.750	6.249	84.354			
5	.560	4.663	89.017			
6	.411	3.425	92.442			
7	.341	2.842	95.283			
8	.200	1.666	96.949			
9	.128	1.063	98.012			
10	.106	.882	98.894			
11	7.498E-02	.625	99.519			
12	5.773E-02	.481	100.000			

Extraction Method: Principal Component Analysis.

a. Only cases for which Class = 13 are used in the analysis phase.

Component Matrix^{a,b}

	Component	
	1	2
Landpct	.852	-.144
CACoV	.909	8.812E-02
CAD	.934	.120
CLAND	.822	.211
ED	.500	-.311
IJI	.220	-.758
LPI	.802	.169
MCA	.790	.161
MNN	-.089	.691
MPI	.903	3.838E-02
MPS	.928	-.012
PSCoV	.877	-.165

Extraction Method: Principal Component Analysis.

- a. 2 components extracted.
- b. Only cases for which Class = 13 are used in the analysis phase.

1.7. Pine

Total Variance Explained^f

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.150	59.584	59.584	7.150	59.584	59.584
2	1.339	11.158	70.742	1.339	11.158	70.742
3	1.044	8.700	79.442	1.044	8.700	79.442
4	.803	6.694	86.136			
5	.573	4.771	90.906			
6	.456	3.801	94.707			
7	.319	2.662	97.369			
8	.118	.983	98.352			
9	8.202E-02	.684	99.036			
10	6.334E-02	.528	99.564			
11	4.459E-02	.372	99.935			
12	7.780E-03	6.483E-02	100.000			

Extraction Method: Principal Component Analysis.

- a. Only cases for which Class = 14 are used in the analysis phase.

Component Matrix^{a,b}

	Component		
	1	2	3
Landpct	.892	-.157	4.350E-02
CACoV	.883	.141	-.008
CAD	.801	.207	-.077
CLAND	.943	-.267	5.004E-02
ED	.563	.475	-.134
IJI	7.277E-02	.709	.527
LPI	.948	-.096	-.023
MCA	.830	-.138	9.653E-02
MNN	-.147	-.277	.850
MPI	.818	-.267	4.020E-02
MPS	.944	-.206	6.032E-02
PSCoV	.743	.484	-.006

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

b. Only cases for which Class = 14 are used in the analysis phase.

1.8. Manuka

Total Variance Explained^a

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.076	50.637	50.637	6.076	50.637	50.637
2	1.521	12.672	63.308	1.521	12.672	63.308
3	1.256	10.466	73.774	1.256	10.466	73.774
4	.865	7.209	80.984			
5	.704	5.865	86.848			
6	.476	3.965	90.813			
7	.401	3.338	94.151			
8	.262	2.182	96.333			
9	.158	1.317	97.649			
10	.124	1.036	98.685			
11	9.599E-02	.800	99.485			
12	6.177E-02	.515	100.000			

Extraction Method: Principal Component Analysis.

a. Only cases for which Class = 15 are used in the analysis phase.

Component Matrix^{a,b}

	Component		
	1	2	3
Landpct	.807	-.361	-.005
CACoV	.883	-.177	-.105
CAD	.797	.209	.246
CLAND	.816	.326	-.027
ED	.228	.500	.706
IJI	-.070	-.471	.587
LPI	.695	.579	5.308E-02
MCA	.766	.175	-.105
MNN	-.312	.482	-.507
MPI	.782	-.276	-.246
MPS	.903	-.160	-.066
PSCoV	.833	-.190	-.072

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

b. Only cases for which Class = 15 are used in the analysis phase.

1.9. Gorse

Total Variance Explained^f

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.604	46.697	46.697	5.604	46.697	46.697
2	1.909	15.907	62.604	1.909	15.907	62.604
3	1.665	13.874	76.479	1.665	13.874	76.479
4	.944	7.868	84.347			
5	.790	6.580	90.926			
6	.342	2.848	93.775			
7	.305	2.543	96.318			
8	.192	1.598	97.915			
9	.136	1.132	99.047			
10	9.531E-02	.794	99.841			
11	1.900E-02	.158	100.000			
12	4.166E-05	3.471E-04	100.000			

Extraction Method: Principal Component Analysis.

a. Only cases for which Class = 19 are used in the analysis phase.

Component Matrix^{a,b}

	Component		
	1	2	3
Landpct	.746	-.213	.373
CACoV	.883	.212	-.318
CAD	.853	.222	-.343
CLAND	.836	.253	-.383
ED	.310	.453	.722
IJI	.188	-.658	.155
LPI	.273	.702	.586
MCA	.858	.230	-.341
MNN	-.197	.338	-.196
MPI	.834	-.187	.205
MPS	.817	-.285	.250
PSCoV	.704	-.539	.116

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

b. Only cases for which Class = 19 are used in the analysis phase.

2. Appendix B: Individual Class Models

Appendix B lists the SAS results from the multivariable regression performed for each vegetation class.

2.1. Prerequisites

```
/*Import Microsoft Access Relation 'tblMultivariateModel' into SAS Library*/
PROC IMPORT OUT= MVS.AllVariables
            DATATABLE= "tblMultivariateModel"
            DBMS=ACCESS2000 REPLACE;
            DATABASE="G:\MVS\Analysis2\ClassAnalysis2.mdb";
RUN;

/*Create temporary formats for categorical variables*/
/*Use under scores to put first category at the top*/
proc format;

            value Quart_f          1='_1 Q1'
                                   2='_2 Q2'
                                   3='_3 Q3'
                                   4='_4 Q4';

            value Thirds_f         2='_2 Q2'
                                   3='_3 Q3'
                                   4='_4 Q4';

            value Double_f         0='_0'
                                   1='_1';

run;
```

2.2. Podocarp / Broadleaf

2.2.1. SAS Code

```
proc genmod data = MVS.AllVariables;
  class class2MPS_T;
  model CountYearsPositive/YearsBeforeControl = Class2IJI_10 Class2LANDpct Class2MPS_T
Class2ED / D=B Type3 LRCI;
  format class2MPS_T thirds_f.;
run;
```

```
proc genmod data = MVS.AllVariables;
  class class2MPS_T;
  model CountYearsPositive/YearsBeforeControl = Class2IJI_10 Class2LANDpct Class2MPS_T
Class2ED / D=B Type3 LRCI;
  output out=a pred=phat;
data b;
  set a;
  w = phat*(1-phat);
proc reg data = b;
  weight w;
  model PercentYearsPositive = Class2IJI_10 Class2LANDpct Class2MPS Class2ED / TOL VIF;
run;
```

2.2.2. Model Result

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio		Chi- Square	Pr > ChiSq
				95% Confidence Limits			
Intercept	1	-1.7561	0.1719	-2.1033	-1.4287	104.39	<.0001
Class2IJI_10	1	0.1467	0.0218	0.1050	0.1904	45.49	<.0001
Class2Landpct	1	0.0116	0.0142	-0.0165	0.0394	0.67	0.4139
Class2MPS_T 3 Q3	1	0.6455	0.0999	0.4501	0.8419	41.73	<.0001
Class2MPS_T 4 Q4	1	0.6547	0.1226	0.4146	0.8953	28.53	<.0001
Class2MPS_T _2 Q2	0	0.0000	0.0000	0.0000	0.0000	.	.
Class2ED	1	-0.0031	0.0006	-0.0042	-0.0021	31.87	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

2.2.3. Model Check

Parameter Estimates

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class2IJI_10	Class2IJI_10	1	0.90411	1.10606
Class2Landpct	Class2Landpct	1	0.39540	2.52908
Class2MPS	Class2MPS	1	0.38673	2.58582
Class2ED	Class2ED	1	0.75601	1.32274

2.3. Broadleaf Scrub / Pure

2.3.1. SAS Code

```

proc genmod data = MVS.AllVariables;
  class class4mnn_T;
  model CountYearsPositive/YearsBeforeControl = Class4IJI_10 Class4LANDpct Class4MNN_T/ D=B
Type3 LRCI;
  format class4MNN_T thirds_f.;
run;

proc genmod data = MVS.AllVariables;
  class class4mnn_T;
  model CountYearsPositive/YearsBeforeControl = Class4IJI_10 Class4LANDpct Class4MNN_T/ D=B
Type3 LRCI;
  output out=a pred=phat;
data b;
  set a;
  w = phat*(1-phat);
proc reg data = b;
  weight w;
  model PercentYearsPositive = Class4IJI_10 Class4LANDpct Class4MNN/ TOL VIF;
run;

```

2.3.2. Model Result

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio	95% Confidence Limits	Chi-Square	Pr > ChiSq	
Intercept	1	-1.3105	0.0816	-1.4726	-1.1526	257.94	<.0001	
Class4IJI_10	1	0.1280	0.0159	0.0969	0.1593	64.56	<.0001	
Class4Landpct	1	0.2115	0.0944	0.0255	0.3963	5.02	0.0251	
Class4MNN_T	2	1	-0.1427	0.1217	-0.3823	0.0951	1.37	0.2411
Class4MNN_T	3	1	0.1334	0.1042	-0.0710	0.3377	1.64	0.2007
Class4MNN_T	4	0	0.0000	0.0000	0.0000	.	.	
Scale	0	1.0000	0.0000	1.0000	1.0000			

2.3.3. Model Check

Parameter Estimates

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class4IJI_10	Class4IJI_10	1	0.95890	1.04287
Class4Landpct	Class4Landpct	1	0.95713	1.04479
Class4MNN	Class4MNN	1	0.95897	1.04278

2.4. Broadleaf Scrub / Manuka

2.4.1. SAS Code

```
proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class5IJI_10 Class5LANDpct class5MPI/ D=B
    Type3 LRCI;
run;

proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class5IJI_10 Class5LANDpct class5MPI/ D=B
    Type3 LRCI;
    output out=a pred=phat;
data b;
    set a;
    w = phat*(1-phat);
proc reg data = b;
    weight w;
    model PercentYearsPositive =Class5IJI_10 Class5LANDpct class5MPI/ TOL VIF;
run;
```

2.4.2. Model Result

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-1.5555	0.1118	-1.7795	-1.3408	193.47	<.0001
Class5IJI_10	1	0.1193	0.0154	0.0894	0.1499	59.80	<.0001
Class5Landpct	1	-0.0552	0.0692	-0.1920	0.0797	0.64	0.4253
Class5MPI	1	0.0764	0.0224	0.0329	0.1211	11.58	0.0007
Scale	0	1.0000	0.0000	1.0000	1.0000		

2.4.3. Model Check

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class5IJI_10	Class5IJI_10	1	0.88763	1.12660
Class5Landpct	Class5Landpct	1	0.56531	1.76894
Class5MPI	Class5MPI	1	0.55362	1.80630

2.5. Pine / Mixed

2.5.1. SAS Code

```

proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class7IJI_10 Class7LANDpct Class7MPS / D=B
    Type3 LRCI;
run;

proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class7IJI_10 Class7LANDpct Class7MPS/ D=B
    Type3 LRCI;
    output out=a pred=phat;
data b;
    set a;
    w = phat*(1-phat);
proc reg data = b;
    weight w;
    model PercentYearsPositive =Class7IJI_10 Class7LANDpct Class7MPS/ TOL VIF;
run;

```

2.5.2. Model Result

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-1.6307	0.1081	-1.8468	-1.4227	227.43	<.0001
Class7IJI_10	1	0.1707	0.0164	0.1390	0.2032	108.64	<.0001
Class7Landpct	1	0.0309	0.0186	-0.0058	0.0673	2.76	0.0968
Class7MPS	1	-0.8708	0.3561	-1.5785	-0.1809	5.98	0.0145
Scale	0	1.0000	0.0000	1.0000	1.0000		

2.5.3. Model Check

Parameter Estimates

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class7IJI_10	Class7IJI_10	1	0.84595	1.18210
Class7Landpct	Class7Landpct	1	0.65432	1.52830
Class7MPS	Class7MPS	1	0.58136	1.72009

2.6. Willows

2.6.1. SAS Code

```
proc genmod data = MVS.AllVariables;
  model CountYearsPositive/YearsBeforeControl = Class9IJI_10 Class9LANDpct Class9LPI class9ED
  / D=B Type3 LRCI;
run;
```

```
proc genmod data = MVS.AllVariables;
  model CountYearsPositive/YearsBeforeControl = Class9IJI_10 Class9LANDpct Class9LPI class9ED
  / D=B Type3 LRCI;
  output out=a pred=phat;
data b;
  set a;
  w = phat*(1-phat);
proc reg data = b;
  weight w;
  model PercentYearsPositive =Class9IJI_10 Class9LANDpct Class9LPI class9ED / TOL VIF;
run;
```

2.6.2. Model Result

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-0.8527	0.1663	-1.1830	-0.5306	26.30	<.0001
Class9IJI_10	1	0.0637	0.0240	0.0171	0.1114	7.03	0.0080
Class9Landpct	1	0.0531	0.0187	0.0163	0.0898	8.04	0.0046
Class9LPI	1	-0.0277	0.0093	-0.0476	-0.0112	8.92	0.0028
Class9ED	1	-0.0008	0.0002	-0.0012	-0.0003	12.17	0.0005
Scale	0	1.0000	0.0000	1.0000	1.0000		

2.6.3. Model Check

Parameter Estimates

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class9IJI_10	Class9IJI_10	1	0.96022	1.04143
Class9Landpct	Class9Landpct	1	0.88684	1.12759
Class9LPI	Class9LPI	1	0.58238	1.71709
Class9ED	Class9ED	1	0.57462	1.74027

2.7. Manuka / Gorse

2.7.1. SAS Code

```

proc genmod data = MVS.AllVariables;
    class class13CAD_D;
    model CountYearsPositive/YearsBeforeControl = Class13IJI_10 Class13LANDpct class13CAD_D /
D=B Type3 LRCI;
    format class13CAD_D double_f.;
run;

proc genmod data = MVS.AllVariables;
    class class13CAD_D;
    model CountYearsPositive/YearsBeforeControl = Class13IJI_10 Class13LANDpct class13CAD_D/
D=B Type3 LRCI;
    output out=a pred=phat;
data b;
    set a;
    w = phat*(1-phat);
proc reg data = b;
    weight w;
    model PercentYearsPositive =Class13IJI_10 Class13LANDpct class13CAD_D/ TOL VIF;
run;

```

2.7.2. Model Result

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-1.4050	0.1096	-1.6232	-1.1934	164.33	<.0001
Class13IJI_10	1	0.1026	0.0185	0.0667	0.1393	30.69	<.0001
Class13Landpct	1	0.0020	0.0250	-0.0472	0.0510	0.01	0.9368
Class13CAD_D	1	1.1008	0.1324	0.8427	1.3619	69.16	<.0001
Class13CAD_D	_0	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000		

2.7.3. Model Check

Parameter Estimates

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class13IJI_10	Class13IJI_10	1	0.92395	1.08230
Class13Landpct	Class13Landpct	1	0.57007	1.75418
Class13CAD_D	Class13CAD_D	1	0.59559	1.67899

2.8. Pine

2.8.1. SAS Code

```

proc genmod data = MVS.AllVariables;
    class class14MNN_T;
    model CountYearsPositive/YearsBeforeControl = Class14IJI_10 Class14LANDpct class14MNN_T /
D=B Type3 LRCI;
    format class14MNN_T thirds_f.;
run;

proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class14IJI_10 Class14LANDpct class14MNN / D=B
Type3 LRCI;
    output out=a pred=phat;
data b;
    set a;
    w = phat*(1-phat);
proc reg data = b;
    weight w;
    model PercentYearsPositive =Class14IJI_10 Class14LANDpct class14MNN/ TOL VIF;
run;

```

2.8.2. Model Result

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Likelihood Ratio		Chi-Square	Pr > ChiSq
				95% Confidence Limits			
Intercept	1	-1.3833	0.1210	-1.6223	-1.1477	130.61	<.0001
Class14IJI_10	1	0.1280	0.0154	0.0980	0.1582	69.43	<.0001
Class14Landpct	1	-0.0022	0.0077	-0.0181	0.0126	0.08	0.7773
Class14MNN_T 3 Q3	1	0.2820	0.1073	0.0715	0.4923	6.90	0.0086
Class14MNN_T 4 Q4	1	0.1400	0.1031	-0.0617	0.3425	1.84	0.1745
Class14MNN_T _2 Q2	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale	0	1.0000	0.0000	1.0000	1.0000		

2.8.3. Model Check

Parameter Estimates

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class14IJI_10	Class14IJI_10	1	0.99812	1.00188
Class14Landpct	Class14Landpct	1	0.98579	1.01442
Class14MNN	Class14MNN	1	0.98609	1.01410

2.9. Manuka

2.9.1. SAS Code

```
proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class15IJI_10 Class15LANDpct class15MPS /
D=B Type3 LRCI;
run;

proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class15IJI_10 Class15LANDpct class15MPS/ D=B
Type3 LRCI;
    output out=a pred=phat;
data b;
    set a;
    w = phat*(1-phat);
proc reg data = b;
    weight w;
    model PercentYearsPositive =Class15IJI_10 Class15LANDpct class15MPS/ TOL VIF;
run;
```

2.9.2. Model Result

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-1.5212	0.1786	-1.8801	-1.1793	72.57	<.0001
Class15IJI_10	1	0.0717	0.0239	0.0257	0.1194	9.02	0.0027
Class15Landpct	1	0.0308	0.0144	0.0025	0.0591	4.55	0.0330
Class15MPS	1	0.7376	0.3286	0.0940	1.3841	5.04	0.0248
Scale	0	1.0000	0.0000	1.0000	1.0000		

2.9.3. Model Check

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class15IJI_10	Class15IJI_10	1	0.99618	1.00383
Class15Landpct	Class15Landpct	1	0.29711	3.36579
Class15MPS	Class15MPS	1	0.29717	3.36503

2.10. Gorse

2.10.1. SAS Code

```
proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class19IJI_10 Class19LANDpct class19LPI / D=B
    Type3 LR CI;
run;

proc genmod data = MVS.AllVariables;
    model CountYearsPositive/YearsBeforeControl = Class19IJI_10 Class19LANDpct class19LPI / D=B
    Type3 LR CI;
    output out=a pred=phat;
data b;
    set a;
    w = phat*(1-phat);
proc reg data = b;
    weight w;
    model PercentYearsPositive =Class19IJI_10 Class19LANDpct class19LPI / TOL VIF;
run;
```

2.10.2. Model Result

Parameter	DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-1.8906	0.2135	-2.3218	-1.4840	78.44	<.0001
Class19IJI_10	1	0.1779	0.0272	0.1260	0.2328	42.80	<.0001
Class19Landpct	1	0.1400	0.0323	0.0764	0.2033	18.77	<.0001
Class19LPI	1	-0.1715	0.0290	-0.2305	-0.1168	34.99	<.0001
Scale	0	1.0000	0.0000	1.0000	1.0000		

2.10.3. Model Check

Variable	Label	DF	Tolerance	Variance Inflation
Intercept	Intercept	1	.	0
Class19IJI_10	Class19IJI_10	1	0.92087	1.08593
Class19Landpct	Class19Landpct	1	0.83800	1.19332
Class19LPI	Class19LPI	1	0.84882	1.178