The use of habitat analysis in

the control of wildlife tuberculosis

in New Zealand

A thesis presented in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy at Massey University

Joanna McKenzie

1999

Abstract

The brushtail possum (*Trichosurus vulpecula*) acts as a wildlife reservoir of *Mycobacterium bovis* infection for farmed livestock and other wildlife species in areas covering an estimated 23 percent of New Zealand. This reservoir of infection is severely hampering efforts to eradicate TB from farmed animals in New Zealand. New strategies need to be developed that focus on identifying areas of residual TB infection in possums and eradicating them, while maintaining low broad-scale population levels in the most cost-effective way. This thesis comprises a number of studies that apply multivariate modelling techniques and spatial analytical procedures to evaluate spatial patterns of possum habitat and of TB occurrence in cattle herds, and hence to identify high-risk localities for possum-derived TB and target control strategies of varying intensity where each will be most cost-effective.

An initial case-control study of 132 possum capture sites was used to identify habitat and topographic factors at the scale of the possum home range that were predictors of the probability of capturing tuberculous possums. Data on the following factors were collected from a 50 meter square area at each capture site: abundance of plant species and structural features of vegetation, the abundance and quality of available den sites, and topographic factors such as slope and aspect. Data were analysed using unconditional logistic regression modelling. Slope of the land and factors associated with the quality of dens were associated with the probability of capturing multiple TB possums. Slope of the land, height of trees, percent cover of vegetation in the height range 0.3 – 2 metres and percent of the ground covered in vascular vegetation were associated with the probability of capturing single TB possums.

A subsequent case control study was conducted at the larger scale of farms to identify possum habitat, geographic and farm factors associated with the probability of tuberculous possums being present on a farm, using random effects logistic regression modelling. A geographic information system (GIS) was used to combine and analyse digital geographic data sets to produce variables for the statistical analysis. Geographic data used for the study comprised farm boundaries, rivers, slope, and vegetation coverage which was derived from satellite imagery. Fragstats, a habitat pattern analysis programme, was used to produce variables representing patterns of possum habitat on farms. Variables were included in the model to account for spatial autocorrelation patterns in the data and a variogram analysis was conducted on residuals to test for any significant spatial autocorrelation that was not

accounted for in the model. Distance from the coastal forest, total area of the farm, area of pine forest, area of manuka/pasture and proportion of the farm covered with podocarp/broadleaved species were the significant variables in the final model.

An automated classification procedure was used to generate a vegetation map from a SPOT3 multi-spectral satellite image for this study. The resulting raster map comprised 8 vegetation classes, including 3 tree classes, 4 scrub classes and a pasture class. This map had a high degree of spatial accuracy and more detailed vegetation classes than existing digital data available in New Zealand. However, some categories contained a heterogeneous range of plant species which limited the discriminatory power of models using this vegetation data.

Spatial clustering of percent TB-positive years, five-year cumulative incidence and space-time clustering of annual cumulative incidence of TB in cattle were analysed for beef breeding farms in the area in which the farm-level case control study was conducted. Analyses used the spatial and space-time scan statistics which are available in SaTScan, a cluster analysis software package. Mapping of farms included in the resulting clusters enabled us to develop hypotheses on the association of different habitat patterns with the resulting disease patterns in cattle. This study also highlighted the value of these cluster analysis techniques as tools in surveillance for wildlife TB, by using farm-level patterns of TB as an indicator of the possible presence of wildlife TB.

The final stage of the project described in this thesis was the development of EpiMAN-TB, a spatial decision support system for the management of wildlife-related TB in New Zealand. Development of this tool was undertaken to assist the application of the research results described in this thesis at the field level. The system combined a relational database, map display tools, spatial analysis tools, simple expert systems and simulation models of the spread of TB in possums at the scale of individual farms and administrative regions. The software comprises five modules set within a user-friendly interface. The modules are: possum TB hot spot prediction, farm TB risk prediction, evaluation of possum control programmes at the farm and the regional level, and an administration module.

Acknowledgements

"Humans need adventure – projects with an uncertain outcome. We have a drive to examine, explore, push the limits of whatever we find ourselves involved in, be it music, physics, art, sport, mathematics or mountaineering."

Graeme Dingle (NZ adventurer), New Zealand Geographic, January-March 1996

Starting out on a PhD journey at the same time as starting the journey of "mother of two" was certainly an adventure with an uncertain outcome. Indeed if we knew the outcome of many of the things that we undertook in life we wouldn't begin them at all, depriving ourselves of an opportunity to grow beyond boundaries we didn't know we had. This has been some journey and it has stretched my boundaries in all ways from the intellectual to the physical to the emotional and the spiritual. I drew strength from and received support from many many people through the course of the past eight years, and I would like to acknowledge these people here.

It is customary to leave the acknowledgement of the support of one's partner in life and/or family to the end. However, in recognition of the enormity of the support involved in helping a mother of young children through a PhD I would like to begin with my husband, Pin. I believe that a good relationship allows each partner to achieve their potential and Pin has done this for me with my PhD. My deepest and most sincere thanks go to Pin for his continued support and for providing an enriching home environment for our kids when I wasn't there. Our two children, Tom and Bridget have only known Mum as doing a PhD (or P.D. as Bridget once called it!). I thank them for being tolerant (most of the time) and for the joy that they show in my finishing. A special thanks to my parents, Don and Ali, who have been very supportive particularly in providing a second home to our kids during the school holidays. Special thanks to my sister and brother-in-law, Fi and Dean, who provided me with a peaceful retreat in the country where I wrote a large part of my thesis. Thanks to all our friends and extended family who have supported us through this long and arduous period, really wondering "why's she doing this?!"

Roger Morris, my chief supervisor, and Dirk Pfeiffer, my secondary supervisor, combine a range of skills and experience, vision and commitment that are unequalled anywhere in the world. I am particularly grateful to Roger for providing me with the opportunity to do a PhD,

Joanna McKenzie   21 May 1999
EpiCentre, Institute of Veterinary and Animal Biosciences,
Massey University, Palmerston North

Table of Contents

List of Figures

List of Tables

INTRODUCTION

The brushtail possum (*Trichosurus vulpecula*) was introduced to New Zealand from Australia on multiple occasions between 1837 and 1940 to establish a fur industry (Pracy, 1962). Possums adapted well to the New Zealand environment and populations have spread from the original colonisation sites to cover most of the country. Possums were first found to be infected with TB on the West Coast of the South Island in 1967 and the species has since been shown to be the most important wildlife vector of TB in New Zealand (Anon, 1986; Morris et al., 1994; Jackson, 1995). It has been estimated that infected possum populations cover approximately 23% of the land area of New Zealand (O'Neil and Pharo, 1995) in areas known as vector risk areas (VRAs). This reservoir of infection is severely hampering efforts to eradicate TB from farmed animals in New Zealand and has resulted in a very expensive TB management programme due to the additional cost of controlling possum populations.

The long term goal of the Animal Health Board (AHB) as stated in the national pest management strategy for the five-year period 1995/96-2000/01, is to "eradicate bovine tuberculosis from New Zealand." However, the authors of the strategy go on to say that "eradication is not a realistic possibility within the term of the strategy with current technology. Therefore, the primary focus of the five year strategy is on the reduction, and where technically feasible, elimination of the transmission of *M. bovis* to and within domestic livestock" (Animal Health Board, 1995). The four objectives for the five-year strategy period are:

1. To reduce the number of infected herds in TB Vector-Free Areas from 0.7 percent to 0.2 percent of the total herds in those areas.

2. To prevent the establishment of new TB Vector Risk Areas and/or the expansion of existing TB Vector Risk Areas into farmland free of TB vectors.

3. To decrease the number of infected herds in TB Vector Risk Areas from 17 percent to 11 percent of the total number of herds in those areas.

4. To encourage individuals to take action against TB on their properties and in their herds.

A major component of the AHB's strategy to achieve the objectives listed above is targeted at increased vector control, with the aim of "managing and eliminating disease risk from wild animals through increased vector control operations." Vector control in New Zealand has become a large business with an increase of expenditure from $18 million in 1995 to a projected $31 million in 1999 (Animal Health Board, 1995). Priorities listed by the AHB in the area of vector control are:

1. Establishment of protection zones to prevent leakage of infected vectors from infected areas.

2. Encouragement and assistance to farmers in infected areas to improve their disease status through:

   • assistance with development and implementation of regional and locally initiated vector control programmes;

   • reclassification of areas as vector risks are reduced;

   • assistance to farmers with high risk herds;

   • self-help programmes;

   • one-on-one programmes.

The major strategy applied to possum control programmes since the mid-1980s has been based on the modelling of Barlow (1991; 1993). This involves intensive initial control of possum populations followed by annual or biennial control to hold the population below a threshold at which TB will decline to extinction, hypothesised to be about 40% of the pre-control density (Livingstone, 1991; Hickling, 1995b). As this strategy has been applied to an increasing number of hectares within VRAs, the incidence of TB in farmed cattle and deer has fallen, and to a lesser extent the incidence of herds quarantined for TB control purposes has also fallen (Livingstone, 1997). Despite the reduction in the incidence of cattle and deer TB reactors, problem herds in which persistent reactors occur remain in areas that have been under a wildlife control programme for many years (G. Atkinson and G. Pannett, AgriQuality New Zealand, pers. comm. 1998). Furthermore, historical experience with possum control in New Zealand has shown that continued control is required to maintain possum populations at a low level to prevent the rebuilding of populations and of TB infection through immigration and reproduction of the local population (Anon, 1986; O'Neil and Pharo, 1995).

Hickling (1995a) noted that eradication of TB from populations of possums may prove more difficult than current epidemiological models predict due to the disease retreating into small persistent foci once populations have been controlled. Hickling and Efford (1996) suggest that the outcomes of the pest management strategy for TB that are based on current models may not be as predicted, due to complexity of the relationship between possum control and TB risk in livestock. The task of eradicating TB from possums is far more challenging than that of reducing the risk of transmission of TB from possums to livestock. Morris and Pfeiffer (1995)

made the statement that "no single measure is going to achieve the desired result in controlling wildlife tuberculosis, and the key to progressive success will lie in tighter and tighter integration of a range of measures to reduce the problem to a point where it is no longer a major concern." As TB incidence declines the remaining foci take on greater importance in relation to achieving the goal of eradication. Successful control during this later stage requires new strategies that focus on identifying and clearing residual infection and preventing the re-introduction of infection into clear areas. There is a need to target control at different levels, and it is important to complement large-scale control with focussed efforts designed to suit the needs of particular problem farms or areas.

To develop these more targeted strategies there is a need for more information on factors associated with the distribution of tuberculous possums, both at the smaller possum home range scale and at the larger farm scale. Such information would facilitate the design of control programmes that target control measures more intensively at probable locations of possum TB clusters, where residual infection may remain following possum population reduction. This would enable:

-- More cost-effective programmes to be implemented during the maintenance phase of control programmes by targeting resources at areas based on possum TB risk. Such information may speed the rate of reduction of infected possums, and consequently of infected livestock within an area.

–   The design of control programmes tailored to the conditions on individual problem farms or problem areas. Such information may increase the probability of eradicating TB from these problem farms or areas.

–   Tailored programmes to prevent the expansion of existing VRAs.

–   Tailored programmes to prevent the introduction of infection into areas from which TB has been eradicated.

Information that is critical to achieving the outcomes described above is the ability to predict the location of persistent foci of TB infection in possums. In Chapter 2 of this thesis we describe an initial study that was conducted to identify habitat and topographic factors associated with the distribution of tuberculous possums at the possum home range scale. Given the association of the distribution of TB in possums with spatially explicit factors, Geographical Information Systems (GIS) are a useful tool for research and management of TB in possums. In Chapter 1 we review the application of GIS and spatial analytical

techniques to health data. The development of a national digital database of farm boundaries, Agribase (Sanson and Pearson, 1997), has made it possible to display TB data in a GIS, to explore spatial patterns of TB at the farm level, and to aggregate geographic risk factors to the farm level. In Chapter 3 we describe a study in which we explored spatial patterns of TB incidence in cattle at the farm level in an area where TB has been endemic in possums since the 1960s. Habitat factors are known to be strongly associated with the density of possums and a digital map of possum habitat is critical to any analyses involving geographic data. Existing digital vegetation data sets were reviewed and a new more detailed vegetation data set was produced using satellite imagery (Chapter 4). This data set plus other digital topographic data were used in an analysis of geographic and management factors at the farm level with the aim of identifying habitat and geographic factors that could be used to predict farms with a high risk of TB possums being present (Chapter 5).

Decision making in relation to the management of wildlife-associated TB is very complex, involving information on cattle TB data, geographic information on farm locations and on possum distribution and possum TB distribution, plus likely response to different control measures, frequency and timing of control. To ensure that the management decisions made are the most cost-effective, all information that is available needs to be considered. Computer-aided decision support can assist the decision process by making information more accessible to managers. Advanced information systems have been developed in the veterinary field (Morris et al., 1993; Sanson, 1993). In this study the EpiMAN(FMD) decision support system was modified and further developed to suit the needs of possum-related decision making in New Zealand (Chapter 6). EpiMAN(TB) provides a framework for applying current knowledge on the epidemiology of TB in possums through expert systems, models of the spread of TB in possums at both farm and regional scales and geographic data to possum management decisions.

We present a final discussion of the results of all studies described in this thesis and conclude with a discussion of the application of these results to vector-associated TB control in New Zealand in the General Discussion.

This thesis has been written in the form of a series of papers that have been prepared for publication. Some papers have been submitted for publication at the time of publishing this thesis, in which case the journal to which the paper has been submitted is given on the title page of the relevant chapter. The remaining papers will be submitted following the submission of this thesis.

Bibliography

Animal Health Board. (1995). National Tb Strategy: proposed national pest management strategy for bovine tuberculosis. Animal Health Board, Wellington, New Zealand.

Anon. (1986). History of Tb control scheme. In: S. Hennessy (Ed), Surveillance 13 (3): 4-8.

Barlow, N.D. (1991). A spatially aggregated disease/host model for bovine Tb in New Zealand possum populations. Journal of Applied Ecology 28: 777-93.

Barlow, N.D. (1993). A model for the spread of bovine Tb in New Zealand possum populations. Journal of Applied Ecology 30 (1): 156-64.

Hickling, G. (1995a). Clustering of tuberculosis infection in brushtail possum populations: implications for epidemiological simulation models. In: F. Griffin, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. University of Otago Press, Dunedin, New Zealand. 174-77.

Hickling, G. (1995b). Wildlife reservoirs of bovine tuberculosis in New Zealand. In: F. Griffin, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. University of Otago Press, Dunedin, New Zealand. 276-79.

Hickling, G.J. and Efford, M.G. (1996). Assessing the risk of bovine tuberculosis infection for New Zealand cattle herds in TB vector risk areas (VRAs). Statistics in Ecology and Environmental Monitoring 2: 79-90.

Jackson, R. (1995). Transmission of tuberculosis (*Mycobacterium bovis*) by possums. Unpublished PhD thesis, Massey University, New Zealand.

Livingstone, P.G. (1991). TB in New Zealand - Where have we reached? In: Proceedings of a Symposium on Tuberculosis. Veterinary Continuing Education Publication No. 132, 113-24.

Livingstone, P.G. (1997). Update on the New Zealand TB situation. In: Proceedings of a Seminar on Possum and Mustelid Control Research. National Possum Control Agencies, Wellington, New Zealand. 17-30.

Morris, R.S. and Pfeiffer, D.U. (1995). Directions and issues in bovine tuberculosis epidemiology and control in New Zealand. New Zealand Veterinary Journal 43 (7): 256-65.

8

Morris, R.S., Pfeiffer, D.U., and Jackson, R. (1994). The epidemiology of *Mycobacterium bovis* infections. Veterinary Microbiology 40 (1): 153-77.

Morris, R.S., Sanson, R.L., McKenzie, J.S., and Marsh, W.E. (1993). Decision support systems in animal health. In: M. Thrushfield (Ed), Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine. 188-99.

O'Neil, B.D. and Pharo, H.J. (1995). The control of bovine tuberculosis in New Zealand. New Zealand Veterinary Journal 43 (7): 249-55.

Pracy, L.T. (1962). Introduction and liberation of the opossum (*Trichosurus vulpecula Kerr*) into New Zealand. New Zealand Forest Service Information Series No. 45, 28pp.

Sanson, R. L. (1993). Development of a decision support system for an animal disease emergency. Massey University, New Zealand.

Sanson, R.L. and Pearson, A.P. (1997). Agribase - A national spatial farm database. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.16.1 - 12.16.3.

CHAPTER 1

Spatial analysis of health data: a literature review

Introduction

A popular historical reference point for spatial epidemiology is Snow's map of cholera incidence around the Broad street pump in Soho in 1855 (Thomas, 1990). There was little development of this field following Pasteur's confirmation of the germ theory, when the medical focus was on aetiological agents. However, the revival of modern medical geography is attributed to the French physician, Jacques May, who re-emphasised the environmental influence by proposing the study of disease ecology in the middle of the 20th century, and received acclaim for his investigation of deficiency diseases in the developing world (Thomas, 1990). A number of authors have reviewed the development of the spatial epidemiology field (Verhasselt, 1993; Kitron, 1998). Others have focussed their reviews more strongly on statistical analytical techniques for spatial data (Gesler, 1986; Marshall, 1991; Pfeiffer and Morris, 1994; Gatrell et al., 1995; Gatrell and Bailey, 1996; Pfeiffer, 1996). Several reference books with details of spatial analytical techniques have also been published (Thomas, 1990; Elliott et al., 1993; Fotheringham and Rogerson, 1993b; Bailey and Gatrell, 1995; Arlinghaus, 1996). These publications are evidence of the growing awareness amongst epidemiologists of the importance of including spatial data in analyses. In their introduction Fotheringham and Rogerson, (1993b) make the point that:

> "..space must ultimately be viewed as more than a container of locations devoid of situational context."

Access to computing power has grown through the 1990s with the rapid increase in the power and capacity of personal computers, enabling larger data sets to be handled, and more complex analyses to be run in a short time. At the same time there have been significant developments in the software to handle geographical data. The availability of more simplified, menu-driven geographic information systems (GISs) has dramatically increased the GIS user group beyond the specialist computer-cum-geographic scientist. The explosion of the internet in the mid 1990s has further enhanced the development of this field through the greater exchange of information between veterinary and medical epidemiologists and geographers. Advances in the satellite industry have made extensive digital land cover data sets readily available for all areas of the world, and have made possible the accurate recording of digital geographical coordinates with Global Positioning Systems (GPS).

Spatial data, also referred to as geo-referenced or geographically referenced data, has been defined as observations for which absolute location and/or relative positioning (spatial

arrangement) is taken into account (Anselin, 1992). The analysis of spatial data can be divided into two principal components: classical spatial analysis and spatial data analysis (spatial statistics) (Griffith, 1996). The former has evolved with the science of geography and is referred to by Bailey (1994) as "a general ability to manipulate spatial data into different forms and extract additional meaning as a result." Griffith (1996) describes classical spatial analysis as involving processes such as "spatial operations research (minimum route selection), logical overlaying (identifying areal units possessing joint categorical attributes), triangulated irregular networking (TIN), and buffering (distance bands around points or lines), among others." GISs constitute a powerful technology that has been developed to manage geographically referenced data. They handle very well the tasks involved in classical spatial analysis.

Spatial data analysis is defined as the statistical description and modeling of spatial data (Bailey and Gatrell, 1995). These authors further divide the area of spatial data analysis into three broad categories: data visualisation, exploratory data analysis, and the development of statistical models, although they make the point that the boundaries between the categories are not distinct.

In this review I briefly discuss GISs and some of the more commonly used analytical tools that they provide for handling spatial data related to health events, plus their application as a simulation modelling tool and their application in decision support systems. I briefly discuss the availability of spatial data sets and issues associated with the visualisation of spatial data. The main focus of the review is on spatial data analysis techniques, with emphasis on examples of application of the various techniques in the literature. I do not describe the different analytical techniques in detail but discuss issues associated with the application of some of these techniques, particularly those that are applied in the studies described in this thesis.

Geographic Information Systems

GISs are a computerised set of tools for the collection, editing, storage, integration, analysis and spatial display of spatially referenced data (Maguire, 1991; Burrough, 1986). In essence, GISs are information management systems that combine computerised mapping technology and database management systems. They include spatial data that describes the location of geographical features, and descriptive data that describes the quantitative and/or qualitative characteristics of the spatial features. There have been a number of reviews of the application

of GISs in both the human (Twigg, 1990; de Savigny and Wijeyaratne, 1995; Kitron, 1998) and animal health fields (Sanson et al., 1991; McKenzie, 1996). Reader (1995a) provided a useful review of GIS resources available to epidemiologists in the health field. The many functions of GISs are dealt with in specialised texts on this subject (Burrough, 1986; Maguire et al., 1991), and a detailed description of these is beyond the scope of this paper. I will give a brief summary of the analytical features of GISs that have commonly been applied to health studies with examples of these from the literature.

The most common use of GISs in the health field has been as a cartographic tool to produce maps (Kitron, 1998). In addition the spatial analytical techniques available in GISs can be used to extract data on factors of interest to test hypotheses, using techniques such as overlaying, neighbourhood analysis, buffering, measuring areas and distances. Gatrell et al. (1995) note that GISs are especially powerful when there is more than one database to be explored, and a major strength of a GIS is its *overlay* capability. Data are generally stored in a GIS in separate layers (or coverages) representing different map themes. Layers can be used to create composite maps by overlapping them on a computer display in the same way that clear acetate sheets are overlaid on an overhead projector. This feature can be used to extract new data on composite variables, or to compare the distribution of disease incidence and other environmental variables which can help identify factors that may explain the spatial and temporal distribution of diseases.

The *buffering* capability can be used to create zones of different size around geographical features, such as point locations of disease cases, farms or rivers. Twigg (1990) described the benefit of intersecting disease incidence data with buffers around hazardous sites. Marsh et al. (1991) used buffering to determine the density of swine herds within a 5 kilometer radius of herds infected with Aujeszky's disease and uninfected control herds. The *neighbourhood analysis* capability within a GIS allows an investigator to find and list all features which meet certain criteria and are adjacent to a particular feature. This function can be used to identify the neighbours of study farms, which can then be used to obtain a measure of a neighbourhood variable. *Measuring distances* from risk factors is a capability often used in spatial analyses. When dealing with an areal unit such as a farm, rather than a point location unit, one needs to consider how the distances will be measured. Whether the distance is measured from the nearest boundary or from the centroid of a farm will depend on how the variable is thought to have its effect. *Measuring areas* of different units, for example vegetation types, is another useful function of a GIS. *Network analysis* is a system of

optimising routes and the allocation of resource flows through a system connected by a set of linear features. Essentially, it allows the modelling of networks. This technique was used to develop alternate measures of family planning accessibility in Thailand, taking into consideration different travel times along roads and tracks of varying quality (Entwisle et al., 1997).

Spatial pattern software packages are also available, enabling epidemiologists to represent more detailed features of habitat mosaics in analyses. Fragstats is a spatial pattern analysis programme for quantifying habitat structure (McGarigal and Marks, 1995).

De Savigny et al (1995) conducted a review of progress in the application of GIS to health issues in developing countries in 1991, and found that despite rapid and productive adoption of this tool by sectors such as agriculture, natural resources, demography, urban and regional planning, the health sector had not yet begun to explore the potential utility of GIS for either health research or for health programming. Likewise, Twigg (1990) described the lack of application of GIS by the Department of Health in Britain. Reader (1995b) commented that "there is a growing contingent of mature GIS application areas which possess well-developed spatial databases and a body of experienced users. On the other hand, a belief exists that the typical use of GIS has not progressed far beyond the use of mapping, query, and spatial data inventory management, and that the potential analytical power of the technology to help solve complex societal and environmental problems has yet to be realised." As recently as 1996, Gatrell and Bailey (1996) described the use of GIS in health research as still in its infancy. These authors and others have referred to the fact that GISs have typically been found wanting when it comes to statistical analysis of spatial data (Anselin, 1992; Fotheringham and Rogerson, 1993a; Bailey, 1994; Bailey and Gatrell, 1995; Arlinghaus, 1996; Gatrell and Bailey, 1996).

Spatial data sets

A major factor limiting the application of GISs to health data has been the lack of accurate spatial data sets. Twigg (1990) attributed the lack of spatial analysis of routinely collected health data in the United Kingdom to the lack of spatial detail and spatial consistency between the various data sets collected routinely by the Department of Health. Very often researchers are in the position of having to use disease incidence data that was collected for administrative purposes other than that for which the study was designed. As a result the data is often aggregated to a level that is too general for significant spatial patterns to be identified,

and lacks sufficient detail and accuracy in addresses to enable accurate geo-coding of records. Mott et al. (1995) refer to the "lack of map boundary data and the intrinsic unreliability of other databases" as constraints to the application of GISs to studying the epidemiology of parasitic zoonoses.

Digital maps of the boundaries of administrative areas are now available in many countries of the world, and advances have been made in techniques to geo-code health data to these areas. Gatrell et al. (1995) discuss the use of postcodes to allocate patient records to specific geographical areas. In Britain, a machine-readable file, known as the Central Postcode Directory, links postcodes to both an Ordnance Survey grid reference and an electoral ward. Levine et al. (1995a) describe the process used to geo-reference motor vehicle accidents in Honolulu to census blocks using a standardised dictionary of street names and the Topologically Integrated Geographic Encoding and Reference (TIGER) system available in the USA. Rushton and Lolonis (1996) discuss different methods to assign addresses on birth records to geographical areas that were meaningful in an analysis of birth defect rates in an urban population in the USA. In a study of lead risk associated with houses in South Australia, O'Dwyer (1998) linked the Digital Cadastral Database (DCDB) to property valuation assessments for rating and taxing purposes using parcel identification numbers. In the area of animal health, a map of farms is an important spatial data set. In New Zealand, AgriQuality New Zealand (formerly MAF Quality Management) has been developing such a database, known as Agribase, by attaching an unique farm identification number to land parcels that are managed as one farm (Sanson and Pearson, 1997). In Britain, the Ordnance Survey grid reference of farms is included in the National Tuberculosis Cattle database, facilitating the mapping and spatial analysis of TB data (Clifton-Hadley, 1993). The boundaries of administrative areas may not be the most appropriate for the disease being studied, and efforts are being put into the development of other more appropriate spatial databases in cases where funding can be obtained. For example, studies relating enteric diseases such as giardiasis and campylobacter in humans to water quality in New Zealand cities are restricted by the lack of accurate data on the distribution zones of different water sources within cities (Duncanson et al., 1998; Pirie, 1998). As relevant spatial data sets are built up, applications of GISs will broaden.

The increased availability of remotely sensed data from satellites has provided important spatial databases on vegetation and land use coverage. This is of particular value in diseases that are associated with land cover factors such as vector-borne diseases. A comprehensive

summary of the application of remotely sensed data in studies of vector-borne diseases was provided by Kitron (1998). Extensive coverage of this area can also be found in the special issue of Preventive Veterinary Medicine (Volume 11, Nos 3-4, 1991) that focussed on applications of remote sensing to epidemiology and parasitology.

Global Positioning Systems (GPS) are increasingly being used as a source of spatial data for health studies (Kitron, 1998). These systems use a series of navigational satellites to record the coordinates of a point or line or area in digital format, with a high degree of accuracy and precision. GPSs are very useful in areas where land occupancy can not be linked to a digital cadastral database, particularly in rural areas. Le Sueur et al. (1995) used a handheld GPS to map households as a part of a malaria information system in the KwaZulu region in South Africa. Oranga (1995) discussed plans to use a GPS to collect demographic data that could be used to evaluate the distribution of health facilities in Kenya.

Many spatial data sets only include data for disease cases, with less accurate or no information on the distribution of the population at risk. This lack of information may be handled by randomly allocating the population at risk to locations on the basis of population density within the areal units being studied. When using grid data, controls may be randomly selected from all non-case cells (Pereira and Itami, 1991).

Data visualisation

Maps provide the most succinct summary of descriptive geographical data as the distribution of the characteristic of interest is readily visible to the eye (Elliott et al., 1993). Choropleth maps are the most common way of mapping descriptive (both qualitative and quantitative) disease data. These are maps in which areas of equal value are separated from areas of a different value by boundaries (Burrough, 1986). Care must be exercised in the preparation of such maps, and careful consideration given to the class interval schemes used, the number of classes shown, and the use of particular shading and colour schemes. The inappropriate use of colour combinations and shading patterns results in maps that are tiring to read and difficult to interpret. Walter and Birnie (1991) conducted an extensive review of 49 national, intranational and international health atlases. They provide a very good discussion on factors that influence readers' impression of geographical variation in the data, and propose a set of methodological guidelines for consideration in future atlases as an appendix to the paper. Walter (1993b) discusses the effect of different shading and plotting symbols on the visual perception of map data. Smans and Estève (1993) present some of the ideas on presentation of

disease maps that were developed at the International Agency for Research on Cancer while producing several cancer atlases. They discuss methods associated with risk estimation, categorisation of data, and colour design. Bailey and Gatrell (1995) provide a detailed discussion of different visualisation methods for both point and area data. These authors refer to cartograms as a visualisation method that addresses the issue of the dominance of physically large areas on the visual perception of maps. A useful ground rule for any graphic output is that if the main point of the map can not be inferred within thirty seconds then it is likely to be too complicated or incorrectly designed, and has therefore lost its purpose of getting across information in an abbreviated and easily assimilated form (Paterson, 1995).

Other symbols used to represent magnitude of disease on maps are shapes, particularly circles, using variation in diameter to represent variation in magnitude of disease (Baumann et al, 1994; Kitron and Kazmierczak, 1997; Kitron, 1998). Care is required in the interpretation of these symbols as the human brain perceives variation in size of shapes as variation in the area, not variation in a linear dimension such as width. For example, a circle twice the diameter of another has an area four times greater, and thus the perception of difference in size is markedly distorted (Paterson, 1995). Hungerford (1991b) used bar charts in conjunction with a boundary map to illustrate the incidence of bovine anaplasmosis by state within the USA.

Dot density maps is another visualisation technique that has been used (Hirschmann, 1994). Dots of equal size represent the same number of cases, and are distributed indiscriminately within the area of interest such as a state or a farm. They provide no information on the distribution of disease within each area. The disadvantage of this method is that the density of dots is affected both by the magnitude of the disease being mapped and by the size of the area of interest. This can be misleading by indicating a worse situation in a small area where a smaller number of dots will appear more dense than a larger number of dots in a larger area, where the disease is in fact more prevalent.

Other methods are available for representing the distribution of animal health and related information collected at points in space rather than areas. These include creation of a three dimensional surface using the level of disease as the third dimension, and creation of a contour or isobar map in which lines connect points of equal value. Both methods involve the interpolation of geographical data using the values at each point to create intervening values. The resulting surface will be a more accurate representation of reality with a greater density of sample points. Pfeiffer (1994) used a three dimensional model to represent the distribution of

TB reactors throughout New Zealand in 1992. Pfeiffer (1994) also used contour lines to spatially represent the number of possums trapped at each site in a research area and the incidence of TB in possums trapped at each site. The two maps were compared to identify if the distribution of each factor was the same to help develop hypotheses about the distribution of TB in possums. Laut (1986) used contour lines to spatially represent the degree of difficulty imposed by the landscape on livestock mustering in Northern Australia for TB control purposes. The disadvantage with these two techniques is that they require some knowledge of the underlying geographical boundaries to put the disease distribution in context. The advantage is that they can be used to visually represent the pattern in space of a set of point data.

Statistical data analysis

Several reference books that provide a comprehensive discussion of statistical techniques for analysing spatial data have been published (Thomas, 1990; Elliott et al., 1993; Fotheringham and Rogerson, 1993b; Bailey and Gatrell, 1995; Arlinghaus, 1996). Several papers reviewing these techniques have also been published (Marshall, 1991; Fotheringham and Rogerson, 1993a; Pfeiffer and Morris, 1994; Gatrell et al., 1995; Gatrell and Bailey, 1996; Pfeiffer, 1996; Kitron, 1998). Spatial data are generally represented either as discrete points in space, in which case analyses are conducted on the grid references themselves, or as an aggregation of individual events (such as counts of the number of cases) within fixed areal units (Gatrell et al., 1995). Areal units may be farms, counties, states, countries, or many other configurations, depending on the scale at which the analysis is conducted. The different types of data and the advantages and disadvantages of each are described in more detail in the publications cited at the beginning of this section. Point data generally represent features that are points in real life (true point data); for example, house locations, intensive farming units such as small pig and poultry farms. Points may also be used to represent the location of areas such as farms or counties, often as the centroid of such an area. Care is needed when points are used to represent areas in an analysis as they can significantly distort the spatial characteristics of the objects that they represent (Mackereth, 1998).

Aggregation of data into areas is an important issue which has been discussed by the authors cited at the beginning of this section. Arlinghaus (1996) dedicates a chapter to the discussion of this issue. To obtain an underlying distribution of population at risk, it is often necessary to use area data whose boundaries have no relationship to local health or environment issues

(Diggle et al., 1990; Gatrell et al., 1995). The most commonly available area maps are of administrative areas developed for government departments, such as regions, territorial local authorities, health districts, census blocks, etc. Brown (1996) makes the point that "aggregation is often necessary to uncover pattern in detail or integrate data from various sources." He provides a very useful discussion on aggregation of contiguous areas using a univariate classification procedure to identify regions of low internal variance. Brown examined three possible regionalisations on the basis of aggregated communes in Rwanda: prefectures which are sub-national administrative units, farming system regions which were defined on the basis of similar agricultural activities and cropping systems, and agroclimatic zones which were defined on the basis of environmental variability. Each regionalisation was constructed by aggregating communes into 8-11 regions, and the internal variation of each was measured using the average coefficient of variation within each. Carvalho et al. (1996) discuss an algorithm that they used to aggregate small contiguous geographical areas with similar risk profiles, until an estimated population size was achieved. The risk profile was classified using principal components analysis, followed by a cluster analysis of social-economic indicators. Bailey and Gatrell (1995) discuss methods by which areal units from different maps may be aggregated.

An important issue associated with analysis of area data is described as the modifiable areal unit problem, which refers to the fact that map pattern can vary dramatically if the size and shape of the areal units forming the study region altered (Thomas, 1990; Bailey and Gatrell, 1995). Quite different analytical results can be obtained by changing the configuration of the areal units. Another important issue is the scale at which an analysis is conducted, as different processes may operate at different levels of geographical aggregation. It may be possible to identify specific non-random patterns at a local level, which appear as random variation at a national level. Wilson (1998) discusses the effect of scale on the interpretation of tick distribution in the USA. Fotheringham and Rogerson (1993a) discuss the variability amongst parameters in a multiple linear regression model in terms of their sensitivity to scale. Getis and Franklin (1987) used second order analysis to show that heterogeneity of ponderosa pine trees within a forest was a function of the scale of analysis.

To obtain the greatest spatial resolution of data, it is important to use the smallest available areal units with homogeneous populations. However, this can lead to the problems associated with exploring patterns in small area data as discussed by many authors cited at the beginning of this section. Because the populations forming the denominators, and the counts of cases

themselves, will often be small, disease rates may be highly variable or unstable. The use of 'empirical Bayes estimates' is becoming a popular method to deal with this issue (Langford, 1994). In essence, the method estimates the true incidence of an area as a weighted sum of both the observed incidence rate in the area and the mean rate in all areas (or a local neighbourhood of the zone in question). Weighting is based on the population within each area, such that incidence values for large populations are given a high weight as the estimate of disease risk is more reliable, and that for small populations is given less weight and is pulled closer to the global or neighbourhood mean.

Exploratory methods

Exploratory methods of spatial data analysis summarise and describe map pattern and relationships within and between maps (Gatrell and Bailey, 1996). The choice of analytical method depends on whether area or point data are available. In the case of point data analyses, the focus of interest is on the distribution of the event locations themselves; for example, distribution of disease cases or controls. The focus of spatial analytical techniques for area data is on the spatial distribution of attributes of the areas studied; for example, disease incidence rates (Bailey and Gatrell, 1995). The main aim of exploratory analyses on both types of spatial data is to determine if there is significant clustering of the data in both space and in time and space. This can give valuable insights into the disease process underlying the pattern and also on possible risk factors associated with the disease. There are a number of different techniques available for testing for clustering in both point data and area data.

Clustering of point data

Kulldorff et al. (1997) give a very useful brief summary of the statistical tests available for spatial and space-time clustering in the Help file for the SaTScan software, with references to papers describing the different techniques. Kulldorff has written a more detailed review paper describing the appropriate use of the different spatial and space-time cluster detection methods in (Kulldorff, 1998). Cluster analyses may be divided into the following five categories: descriptive cluster detection methods without hypothesis testing, cluster detection tests that also identify the location and test the significance of clusters, focused tests to detect an elevated disease risk around a specific source, global space clustering tests and space-time interaction tests (Kulldorff et al., 1997).

Openshaw et al.'s (1987) geographic analysis machine (GAM) and Rushton's D-map (URL) provide techniques for detecting clusters by using a moving circle of a fixed diameter that

centres on points at regular intervals within an area of interest. The methods compare the disease risk within the circle to the disease risk of the whole population. Rushton and Lolonis (1996) describe the application of their technique to detect clustering of birth defect rates. A criticism of these methods is that a separate significance test is made for each circle, giving rise to a problem with multiple testing, which can lead to the possible detection of a number of false positive clusters that have occurred due to chance. Kulldorff et al. (1997) note that these methods are both very useful for descriptive purposes, but should not be used for hypothesis testing.

The spatial scan statistic available within SaTScan (Kulldorff et al., 1997) is a cluster detection test that is able to pinpoint the location of specific clusters and test the significance of clusters. This method also uses a moving circular window to detect clusters. However, in the case of the spatial scan statistic the diameter of the circle varies in size at each point location up to a maximum set in the analysis. This has the advantage of being useful in a data set that contains points with a variable density, and also overcomes the problem of pre-selection bias when testing for purported clusters as it tests for clustering within areas of all sizes at each location. The method compares the risk of disease within each circle to that outside the circle. It uses a hypothesis testing technique that overcomes the multiple testing problem (Kulldorff and Nagarwalla, 1995). A further advantage of this method is that it it lists the geographic coordinates of the points included within significant clusters, allowing the location of clusters to be mapped. The spatial scan statistic has been used to investigate clustering of childhood leukaemia in Sweden (Hjalmars et al., 1996) and of breast cancer in the northeast United States (Kulldorff et al., 1997).

Focused tests are used to test if there is an elevated risk of disease around a pre-specified point source. The SaTScan program can be used to conduct a focused test, and other tests listed by Kulldorff et al. (1997) include Bithell's test, the Lawson-Waller score test, Stone's test and isotonic binary regression. Kulldorff warns that focused tests should never be used when the point source was defined using the data itself, due to the effect of pre-selection bias.

Most of the available spatial clustering techniques test for global clustering; that is, they indicate whether there is evidence of spatial clustering in the data set or not, without specifying which points are included in the cluster(s). For example, the Cuzick and Edwards test and K nearest neighbour analysis were used to test for clustering of papillomatous digital dermatitis in dairy herds in California (Rodríguez-Lainz et al., 1996), and of bluetongue in cattle herds in Queensland (Ward et al., 1996). The Cuzick and Edwards test and nearest

neighbour analysis were used to test for clustering of turkey farms with fowl cholera in California (Carpenter et al., 1994). Hungerford (1991a) used second order analysis of points to test for clustering of herds with anaplasmosis within different breed groups, using a Monte Carlo technique to test for the significance of different levels of clustering in the different groups. Gatrell et al. (1995) briefly describe the application of Diggle and Chetwynd's K function to test for spatial clustering of point data controlling for the distribution of the population at risk.

Testing for the presence of clustering in a combination of space and time can also provide useful insights into the mechanisms associated with disease transmission. A number of different tests have been developed to test for space-time clustering. These methods are designed to evaluate whether cases that are close in space are also close in time and vice versa, adjusting for any purely spatial or purely temporal clustering (Kulldorff et al., 1997). Knox's test for space-time clustering is a useful test for categorical data (Carpenter et al., 1994; Paré et al., 1996), while Mantel's test is useful for numerical data (Bailey and Gatrell, 1995). These techniques test for the presence of global space-time interaction; that is, they indicate whether there is significant clustering throughout the study region and time period without pinpointing the location of the space-time clusters, either in space or in time. The space-time scan statistic available within the SaTScan program is able to both test for the presence of significant space-time clustering and to pinpoint the location of clusters in space and in time (Kulldorff et al., 1998).

Examples of other techniques for the analysis of spatial patterns in point data are available in the literature. Kitron et al. (1991) compared the mean distance from a national park of deer infested with ticks carrying the bacterial agent causing Lyme disease with those carrying uninfested ticks, using Cochrans' modified t-test. Rodríguez-Lainz et al. (1996) compared the proportion of herds with and without papillomatous digital dermatitis in areas of different sizes using the Kolmogorov-Smirnoff two-sample test. Elliott et al. (1994) compared the prevalence of disease in Bighorn sheep with a fixed-size shifting frame analysis using log-likelihood ratio tests.

Descriptive techniques such as mean centre, standard distances, and standard deviational ellipses were used to compare the distribution of different types of motor vehicle accidents in Honolulu (Levine et al., 1995b), and of cattle herds with and without a history of badger-associated tuberculosis in Britain (Clifton-Hadley, 1993). Such techniques are useful as a summary measure with which to compare different sub-groups in the population. However,

they are susceptible to the effects of extreme values, or outliers.

Clustering of area data

The most common techniques used to analyse spatial patterns in area data, focus on spatial autocorrelation, which quantifies the correlation between values of the same attribute at different spatial locations (Bailey and Gatrell, 1995). Significant autocorrelation indicates significant clustering of the attribute value being measured. The two measures that have been most widely used are Moran's I and Geary's C. The general tool that is used in these calculations is a spatial weights matrix W (also referred to as a proximity matrix), with number of rows and columns equal to the number of spatial units in the study. Each cell in the matrix, w_{ij} represents a measure of the spatial proximity of areas A_i and A_j. In general, if two objects are considered to be in close proximity w_{ij} is set to 1, alternatively it is set to zero. Values between 0 and 1 may also be used.

An important issue is how to represent proximity of observations when they may relate to irregularly shaped areal units. We can not simply use distances between the centroids of each of the areas, as we would be disregarding some aspects of the spatial nature of these areas (Bailey and Gatrell, 1995). The way in which spatial proximity between areal units is defined depends on the disease process being modelled and the particular mechanisms through which one believes spatial dependence to arise. Different weight matrices can be used to put emphasis on particular patterns and to test different hypotheses about the relationship between areal units (Hungerford, 1991b).

Arlinghaus (1996) presents an extensive discussion on the identification of spatial weights matrices, with references to other publications on this topic. Bailey and Gatrell (1995) and Walter (1993a) describe a number of options for assessing the spatial proximity of areas, which are listed below.

- A_j shares a common boundary with A_i.

- The centroid of A_j is one of the k nearest centroids to that of A_i

- Centroid of A_j is within some specified distance of that of A_i. For example, Carpenter et al. (1994) used a radius tool to develop a weights matrix to measure autocorrelation between fowl cholera outbreaks on turkey farms in California.

- The proportion of the boundary of A_i that abuts A_j.

- The length of the common boundary between A_i and A_j divided by the perimeter of A_i is

greater than a certain proportion.

- Inverse distance between points A_i and A_j; for example, Kitron and Kazmierczak (1997) used the inverse of the distance between county centroids to assign highest weights to pairs of neighbouring counties, and decreasing weights to associations of counties further apart.

- Inverse of the minimum number of boundaries one must cross to move between A_i and A_j.

- Hybrid measures based on these various criteria can also be used; for example, combinations of the length of shared boundary and distance between centroids. For example, Brown (1996) used a weights matrix of $w_{ij} = l_{ij}/d_{ij}^3$ where l equals the length of common border between areas and d the distance between centroids.

We could also use proximity criteria that incorporate various alternative measures of spatial separation (Bailey and Gatrell, 1995). These might include travel time between area centroids, for example. The authors note that there is no reason why W need be symmetric. Hungerford (1991a) used different weight matrices when calculating Moran's I to test different hypotheses regarding factors associated with the spatial distribution of the prevalence of bovine anaplasmosis in counties in Illinois. In addition to using adjacency as a measure of proximity, she used the criterion that counties shared the same cattle market but did not share a common boundary, to isolate the effect of cattle movement on the spatial pattern of anaplasmosis at the county level.

Mackereth (1998) investigated the effect that representing area data as point data, based on the centroids of areas, had on measures of autocorrelation. He found that point data did not represent the proximity of observations as accurately as area data. A Moran's I test for autocorrelation in cattle tuberculosis rates on farms was significant at 100 meters for area data (using the true farm boundaries), but not for point data using the centroids of the same farms. Farms tended to be clustered by farm size, hence the scale at which autocorrelation occurred between centroids of large farms was different to that between centroids of small farms. Bailey and Gatrell (1995) discuss the effect of non-stationarity on autocorrelation analyses, and make the point that analyses can become complex when representing non-stationarity in spatial data. Kitron and Kazmierczak (1997) subdivided the counties of Wisconsin into 3 regions and studied spatial clustering patterns across the state as a whole, and within each region. They found peak clustering of human cases of Lyme disease occurred at different distances in each of the regions. Such regionalisation enables epidemiologists to detect more

detailed patterns at the regional level compared with an average across the whole area. Walter (1993a) examined the power of both Moran's I and Geary's C to detect clustering under different spatial configurations of data. He found that the power of these tests was reduced when high-rate polygons were clustered on the edge of the study area. He also discussed the effect of data quality on the ability to detect spatial patterns.

By testing for autocorrelation across different neighbourhood distances, referred to as different spatial lags, one can generate a spatial correlogram which can be used to identify the distance at which the most significant autocorrelation occurs, i.e. the spatial range over which the spatial clustering effects are maximised. Kitron and Kazmierczak (1997) evaluated Moran's I at 5-kilometer intervals from the centroids of counties for tick distribution, Lyme disease cases, population density and NDVI values.

Other tests are available for testing the spatial patterns between maps. For example, Hungerford (1991a) describes using the kappa statistic test for the overlap of areas on different maps, and another for the co-distribution of two variables at each point using Tj ll stheim's statistic (Tjøstheim, 1978). Jacquez (1995) presents an example of a slightly different variation on this theme with a test for the overlap of map boundaries.

Statistical modelling

Statistical modelling of data involves techniques that aim to explain or estimate the relationship between a dependent variable and a group of independent variables or covariates; for example, between some measure of disease incidence and social and/or environmental covariates. Many approaches have been used for the statistical modelling of spatial data. Buckland and Elston (1993) provide a useful summary of different statistical modelling methods applied to the spatial distribution of wildlife. A number of different predictive modelling approaches for spatial data were compared by Williams et al. (1994). They used linear and non-linear discriminant analysis, tree-based induction and neural networks to map tsetse fly distributions in Zimbabwe and concluded that while the simpler methods (linear discriminant analysis and tree-based induction) were less precise, they were easier to interpret. The most common statistical modelling methods used in the health field are multiple linear regression and logistic regression. Pfeiffer (1996) notes that in the field of epidemiology parameters of interest are very often counts or proportions which can be modelled using generalised linear modelling techniques rather than ordinary least squares regression. A number of authors have included spatial data into multivariate analyses as

independent variables, but have not accounted for spatial dependence in these models. Pfeiffer (1994) used specific geographical variables such as height above sea level, aspect, slope and distance to features of interest as explanatory variables in a multivariate statistical model of the distribution of dens used by tuberculous possums. Glass et al. (1995) developed a risk density map for Lyme disease based on a multiple logistic regression model.

For many years researchers have pointed out that we cannot directly apply these classical statistical models to spatial data as they are based on the assumptions that observations are independent and all observations are drawn from populations with equal variance, and such assumptions are often violated in the case of spatial data (Cook and Pocock, 1983). However, in more recent years methods of representing spatial patterns in statistical models have been explored in more depth, and there is more published literature describing these techniques (Bailey and Gatrell, 1995; Arlinghaus, 1996). There are still relatively few examples in the literature of the incorporation of spatial patterns into regression models. In the case of spatial data, the assumption of independence of observations is often not met, as events at one location are likely to be correlated with events at neighbouring locations (spatial autocorrelation). Cook and Pocock (1983) point out that as long as it is possible to explain the non-independence in terms of covariates that are also spatially correlated, no problem arises. However, unexplained variation in spatial data will often lead to the error term in any multiple regression model being spatially correlated. Bailey and Gatrell (1995) state that the standard regression assumptions that the error term is a vector of independent random errors with constant variance are very unlikely to be appropriate for spatial data because of the possibility of spatial dependence between the residuals. With most real area data this assumption is often violated and residuals from such a regression will be spatially correlated. In addition, it is unlikely that the variance will be constant across all residuals. As a result of one or both of these problems the conventional confidence intervals for regression coefficients and the corresponding assessment of the significance of any of the covariates may be unreliable. This results in an increased probability of rejecting the null hypothesis when it is in fact true (Type I error) as the standard error of parameter estimates tend to be smaller if spatial autocorrelation is not accounted for in the model.

The most commonly recommended procedure to manage the lack of independence of spatial data is to first conduct an ordinary regression analysis without allowing for spatial autocorrelation. The residuals from the initial analysis are then tested for the presence of spatial autocorrelation using a test such as Moran's I. If significant autocorrelation is present,

then variables representing the spatial pattern in the residuals can be identified and included in the original model. This can be considered an iterative process of testing for autocorrelation in residuals and adding spatial pattern variables to the model until there is no significant autocorrelation remaining, indicating that the model is a good representation of the data. Bailey and Gatrell (1995) discuss a number of different classes of *autoregressive models*. Such models include as covariates one or more variables that represent the spatial dependence of the outcome variable.

The challenge for the epidemiologist is finding appropriate variables to represent spatial patterns in the data. One approach is to model the pattern of autocorrelation in the residuals using a variogram. Such a method could be applied to area data using the centroid of areas as point data (Bailey and Gatrell, (1995). The variogram fits a smooth function of distance between point locations that can be incorporated as an independent variable in the final model. Thomson et al. (1998) describe using a variogram of the standardised residuals from the fitted logistic regression model in order to estimate the spatial dependence in their data, in a study to predict the impact of insecticide treated bednets on malaria infection in villages in The Gambia. Based on the appearance of the variogram, they assumed that the correlation between a pair of measurements was an exponentially declining function of distance, i.e. $p(d)$ = exp ($-ad$) for some positive value of a. Such models assume that the spatial autocorrelation effect is stationary across the study area. Bailey and Gatrell (1995) state that "stationarity of the second order component is less likely to hold for area data. Even if there exists an underlying continuous space process that is stationary, indirect observation of this using aggregated values over irregular areas will result in variances and covariances in the areas that will not be the same for all areas. Further, we do not have a simple measure of distance between areas, in that proximity measures other than Euclidean distance between centroids may in many cases be preferable for area data."

Another approach is to try different connectivity matrices that are not necessarily based on Euclidean distance functions between centroids, such as those discussed above under 'Exploratory methods to analyse spatial data' (Bailey and Gatrell, 1995; Arlinghaus, 1996). This approach provides more flexibility in testing different hypotheses regarding spatial dependence in the data. Brown (1996) discusses the sensitivity of spatial autoregressive models to the specification of the weights matrix. He tested different spatial weights matrices to show the sensitivity of the models to this, and found that significant spatial autocorrelation remained in the model residuals when the weights matrix was calculated with the exponent of

distance changed to -1. He concluded that the generalised weights matrix with a proximity function of d^{-3} characterised the spatial interaction in the net migration field better than the other weighting schemes as it rendered a tighter pattern. Thomson et al. (1996) used elements of distance matrices in a multiple linear regression models of bednet usage and malaria prevalence in The Gambia, to account for the confounding effect of the spatial dependence amongst villages included in the study. They did not specify the nature of the distance matrix used. Levine et al. (1995b) included a spatial lag variable in a linear regression model of motor vehicle accidents in Honolulu. These authors discuss the use of Spacestat, a special spatial statistics regression package that estimates linear regression equations controlling for the degree of spatial autocorrelation in the data.

Spatial autocorrelation may be represented by variables other than proximity matrices. For example, Smith et al. (1995) generated a conditional autoregressive model of mosquito density by adding a function that measured the spatially structured variation in the data. The function was the mean mosquito density for all neighbours (within a distance of 150 meters) of each house, not including the house itself. Such models are recommended by Bailey and Gatrell (1995) as they need not assume stationarity of the spatial autocorrelation pattern (the second order effect), nor are they necessarily restricted to covariate structures that are smooth functions of Euclidean distance.

Visual inspection of mapped residuals can provide important information on areas where the model is not fitting the data well. This in turn can help identify variables that represent the different spatial autocorrelation patterns associated with different areas. In a study of factors influencing migration in Rwanda, a map of residuals provided valuable information about controls on migration that were acting at a local level and not nationwide (Brown, 1996). The spatial autocorrelation in the residuals indicated that communes with higher in-migration than that predicted by the model tended to occur in groups in certain parts of the country. Other authors have tested the association of different independent variables (covariates) with the spatial pattern of the outcome variable. Feng (1996) used a linear autoregressive model of regional fertility rates in an area of China to investigate the effect of implicit spatial factors whilst controlling for a number of explicit non-spatial variables. Jacqmin-Gadda et al. (1997) propose a test of the residuals in generalised linear models that is a generalisation of the spatial autocorrelation test based on Moran's I. This allowed adjustment for sizes of geographical areas and for explanatory variables. Langford et al. (1998) used multi-level modelling of geographically aggregated health data to study the effect of UV exposure on

malignant melanoma mortality in the European Community.

Bailey and Gatrell (1995) note that "spatial forms of logistic regression models are not well developed yet." These authors suggest introducing covariates into the regression model such as the spatial coordinates, or a variable representing regions categorised broadly by location to remove the effect of spatial dependence. Pfeiffer et al. (1997) evaluated three different spatial models to predict the occurrence of theileriosis outbreaks in Zimbabwe, using the likelihood ratio to identify that which best explained the spatial pattern of the data. They compared models with spatial patterns represented in the following ways: entering the coordinates of each grid cell included in the analysis; including as a random effect a variable that represented the identity of a 25-cell area surrounding each grid cell, and another that represented a 16-cell area surrounding each grid cell; replacing the binary outcome variable with a binomial variable that represented the probability of theileriosis outbreaks in the immediately neighbouring grid cells; including this same variable as an independent variable with the binary dependent variable. Inclusion of the 25-grid area identification as a random effect variable resulted in the model with the smallest deviance term. The autoregressive model that included the probability of theileriosis in immediately neighbouring grid cells resulted in a model with lower deviance than that without any spatial parameters, while all other variables resulted in models with no significant difference in deviance terms. Augustin et al. (1996) compared the deviance terms of logistic regression models of deer distribution in Scotland that included a spatial function operating over different scales, described as autologistic models. They describe an autocovariance function that weights the probability of deer being present in a cell, as a function of the weighted average of the number of occupied cells in sets of neighbouring cells. They found that the autocovariate corresponding to a neighbouring area of 7 kilometers produced the greatest reduction in deviance.

Spatial filtering methods have also been used to remove spatial dependence from grid data used in multivariate analyses (Kitron, 1998). Pereira and Itami (1991) controlled for spatial autocorrelation in the independent habitat variables used to predict squirrel distribution using a systematic sampling scheme. They detected the maximum distance over which significant spatial autocorrelation occurred using Moran's I, then used this distance to systematically select control cells.

GIS as a simulation modelling tool

GISs have been used in a limited number of simulation models of diseases that are associated with one or more geographical risk factors. Such models represent a dynamic disease situation, and are extremely useful tools to test the effectiveness of different control strategies under different disease scenarios. PossPOP is a simulation model of the spread of TB in possums developed by Pfeiffer (1994). This model imports vegetation data from a GIS to represent the area over which possum den sites are distributed. By basing the model on real geographical data, it can be applied to different farm situations using the actual vegetation cover on those farms, hence providing a more realistic representation of the habitat on the farm, and hopefully more reliable results. The main aim of this model is to compare different possum control strategies, such as frequency and timing of population control, and population control in conjunction with vaccination for TB.

Interspread is a simulation model of the spread of foot-and-mouth disease (FMD) between farms (Sanson et al., 1994). It uses databases containing the actual geographical coordinates of the boundaries of all holdings containing susceptible animals in the study area, as well as the locations of saleyards and other animal congregation sites. The model operates at the farm level, and can be used to test different control strategies. Tinline (1988, cited in Sanson et al., 1991) identified six different factors that affect the persistence of rabies in southern Ontario, using computer simulation modelling.

Models of the spread of airborne disease agents, such as viruses or air pollutants, can be used in a GIS in conjunction with a database of land and occupancy data to identify areas at risk. Windspread is a geographical model of the airborne spread of FMD virus. The model uses a farm as a given starting point for the spread of virus, and uses climatic data, such as wind speed and direction, to simulate the spread. Overlaying the virus plume on a map of farms can be used to identify farms that are at risk of being exposed to the virus, which can help outbreak managers establish priorities for farms requiring surveillance, or the implementation of preventive measures such as vaccination or pre-emptive slaughter. Scott et al. (1998) used a Gaussian plume dispersion model to represent the exposure of dairy and beef herds in Alberta, Canada to sour gas emissions, and associate this with health and productivity factors.

Use of GIS in decision support systems

Decision support systems are useful in situations where decisions need to be made quickly such as in an emergency, and where decisions are complex, requiring information from many sources that needs to be pre-processed before being in a state useful to the decision maker. Such systems can also be programmed to produce maps of disease distribution, updating managers on the progress of control or eradication programs. One such example in the area of animal health is EpiMAN-FMD, a decision support system developed for the management of a FMD outbreak (Sanson, 1993). This decision support system comprises a GIS, a database management system, an expert system and models of the spread of FMD. The GIS includes a map of farm boundaries and topographic information, and interfaces with a farm database which contains livestock numbers, farm owner, address, phone number, and other details. One of the uses of the GIS is to predict windborne spread of FMD virus and to identify farms over which the 'virus plume' has passed and which may be incubating the disease, thus represent high priority for quarantine and monitoring. Another important use of the GIS is to identify and produce a list of farms that are located within distance zones of an infected property and which require quarantine and monitoring.

A similar decision support system is being developed to assist with the management of TB in New Zealand, known as EpiMAN-TB (McKenzie and Morris, 1995; McKenzie et al., 1997). This system also combines a GIS with a farm database, an expert system and models of the spread of TB at the farm scale and the regional scale. EpiMAN-TB can be considered a 'box of tools' with a number of different functions available to TB managers at district, regional and national levels. The 'hot spot predictor' module identifies areas of vegetation with a high risk of TB occurring in the possum population and others with a low risk. The 'on-farm TB control' module uses the TB simulation model, PossPOP, to compare different control strategies at the farm level. The 'farm risk' module will utilise the modelling capability of a GIS and classify farms in an area according to their risk of having a high incidence of TB, a sporadic incidence or no risk of TB, which can be used to target disease surveillance and vector control efforts in newly infected areas.

Laube et al.(1997) describe a GIS-based decision support system for region-wide enzootic pneumonia eradication in pigs. The GIS uses farm ownership data, livestock numbers per farm and locational coordinates, and the simulation runs in ArcInfo (Version 7.0). Running of the GIS-based system in Switzerland has been limited by the lack of an adequate national

farm database, and the authors make the comment that "the full potential of GIS-based decision support systems in the field of animal health will – to a large extent – depend on improvements of the accessibility and compatibility of national farm databases".

Cameron et al. (1997) have developed a computerised Outbreak Response Management System to assist in implementing ring vaccination of cattle and buffalo in the face of an FMD outbreak in Thailand. The system was developed in ArcView using the Avenue programming language. It uses four geographical data sets: village locations, district boundaries, roads and the location of district veterinary offices. The program indicates the number of cattle and buffalo that need to be vaccinated in a buffer zone of specified radius around a specified village where an outbreak has occurred.

Conclusion

In this paper I have aimed to help those new to the field of applying GIS and spatial data analysis techniques to epidemiology understand the tools available to them and some of the issues associated with using spatial data. I have reviewed both the medical and veterinary literature for examples of how these tools have been applied. I have discussed the application of GISs in the visual display of disease data, the preparation/construction of geographic variables for statistical analyses, and the application of geographic data in simulation modelling and in decision support systems. I have described statistical techniques for analysing spatial data, and discussed in some detail different ways of measuring spatial proximity and representing this in spatial data analyses, and in the statistical modelling of spatial data by incorporating variables that represent the spatial patterns in the data.

In 1996 Clarke et al. wrote that,

> "GIS......does not fit neatly into the health scientist's toolbox. It requires rethinking and reorganizing the way that data are collected, used and displayed. It requires expense, training, and a climb up a learning curve. It needs maintenance and support and can be both overwhelming and threatening to the uninitiated."

The application of spatial data into epidemiology is indeed challenging. However, I believe that the number of scientists now working with spatial data in the health field together with the improvements in the user-friendliness of GISs and the greater availability of spatial data is generating a momentum that will help this field continue to grow in the future. As there are

more examples of the application of spatial data in epidemiological studies in the literature, more people are likely to be encouraged to undertake similar studies. This in turn is likely to lead to the availability of a wider range of spatial data and possibly to better quality data as more routine data is collected with the purpose of being used for spatial analyses.

Bibliography

Anselin, L. (1992). Spatial data analysis with GIS: An introduction to application in the social sciences. 75. Technical Report Series. National Center for Geographic Information and Analysis, Santa Barbara, California.

Arlinghaus, S. L., Editor. (1996). Practical handbook of spatial statistics. CRC Press, Boca Raton, Florida, U.S.A.

Augustin, N.H., Mugglestone, M.A., and Buckland, S.T. (1996). An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology 33: 339-47.

Bailey, T. C. (1994). A review of statistical spatial analysis in geographical information systems. In: S. Fotheringham, and P. Rogerson (Eds), Spatial Analysis and GIS. Taylor & Francis, London. 13-44.

Bailey, T. C., and Gatrell, A. C. (1995). Interactive Spatial Data Analysis. Longman Scientific & Technical, Burnt Mill, Harlow, Essex, England.

Brown, D. G. (1996). Spatial statistics and GIS applied to internal migration in Rwanda, Central Africa. In: S. L. Arlinghaus (Ed), Practical handbook of spatial statistics. CRC Press, Boca Raton, Florida, U.S.A. 149-73.

Buckland, S.T. and Elston, D.A. (1993). Empirical models for the spatial distribution of wildlife. Journal of Applied Ecology 30: 478-95.

Burrough, P.A. (1986). Principles of geographical information systems for land resources assessment. Oxford University Press, England.

Cameron, A., Sharma, P., and Chamnanpood, P. (1997). Implementation of a simple outbreak response management system for developing countries using a low-cost GIS. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 2.14.1-2.14.3.

Carpenter, T.E., Snipes, K.P., and Hird, D.W. (1994). A spatial epidemiologic investigation of fowl cholera. In: Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics.

Carvalho, M.S., Cruz, O.G. and Nobre, F.F. (1996). Spatial partitioning using multivariate cluster analysis and a contiguity algorithm. Statistics in Medicine 15 (17-18): 1885-94.

Clifton-Hadley, R.S. (1993). The use of a geographical information system (GIS) in the control and epidemiology of bovine tuberculosis in south-west England. Society for Veterinary Epidemiology and Preventive Medicine. 166-79.

Cook, D.G. and Pocock, S.J. (1983). Multiple regression in geographical mortality studies, with allowance for spatially correlated errors. Biometrics 39: 361-71.

De Savigny, D. and Wijeyaratne, P.E. (1995). GIS for Health and the Environment. In: Proceedings of an International Workshop held in Colombo, Sri Lanka 5-10 September 1994. International Development Research Centre, Ottawa, Canada.

Diggle, P. J., Gatrell, A.C. and Lovett, A.A. (1990). Modelling the prevalence of cancer of the larynx in part of Lancashire: a new methodology for spatial epidemiology. In: R. W. Thomas (Ed), Spatial epidemiology. Pion Limited, London. .35-47.

Duncanson, M., Weinstein, P., Baker, M., Russell, N., Woodward, A., and Skelly, C. (1998). Regional rates of notified cryptosporidiosis compared with drinking-water supply quality in Aotearoa New Zealand. In: P. Firns (Ed), Proceedings of the 10th Annual Colloquium of the Spatial Information Research Centre. University of Otago, Dunedin. 87.

Elliott, L.F., Boyce, W.M., Clark, R.K., and Jessup, D.A. (1994). Geographic analysis of pathogen exposure in bighorn sheep (*Ovis canadensis*). Journal of Wildlife Diseases 30 (3): 315-18.

Elliott, P., J. Cuzick, D. English, and R. Stern, (Eds). (1993). Geographical and Environmental Epidemiology. Methods for Small -Area Studies. World Health Organisation, Oxford, England.

Entwisle, B., R.R. Rindfuss, S.J. Walsh, T.P. Evans, and S.R. Curran. (1997). Geographic information systems, spatial network analysis, and contraceptive choice. Demography 34 (2): 171-87.

Feng, H. M. (1996). Spatial statistical modeling of explicit and implicit determinants of regional fertility rates: a case study of He-Nan Province, China. In: S. L. Arlinghaus (Ed), Practical handbook of spatial statistics. CRC Press, Boca Raton, Florida. 175-229.

Fotheringham, A.S. and Rogerson, P.A. (1993a). GIS and spatial analytical problems. International Journal of Geographical Information Systems 7 (1): 3-19.

Fotheringham, S., and P. Rogerson, (Eds) (1993b). Spatial Analysis and GIS. Taylor & Francis, London.

Gatrell, A.C. and Bailey, T.C. (1996). Interactive spatial data analysis in medical geography. Social Science and Medicine 42 (6): 843-55.

Gatrell, A.C., Collin, J.R., Downes, R., Jones, B. and Bailey, T.C. (1995). The geographical epidemiology of ocular diseases: some principles and methods. Eye 9 (3): 358-64.

Gesler, W. (1986). The uses of spatial analysis in medical geography: a review. Social Science and Medicine 23 (10): 963-73.

Getis, A. and Franklin, J. (1987). Second-order neighbourhood analysis of mapped point patterns. Ecology 68 (3): 473-77.

Glass, G.E., Schwartz, B.S., Morgan, J.M., Johnson, D.T., Noy, P.M., and Israel, E. (1995). Environmental risk factors for Lyme disease identified with geographic information systems. American Journal of Public Health 85 (7): 944-48.

Griffith, D. A. (1996). Introduction: The need for spatial statistics. In: S. L. Arlinghaus (Ed), Practical handbook of spatial statistics. CRC Press, Boca Raton, Florida. 1-30.

Hjalmars, U., Kulldorff, M., Gustafsson, G., and Nagarwalla, N. (1996). Childhood leukaemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. Statistics in Medicine 15: 707-15.

Hungerford, L.L. (1991a). Application of spatial statistics to the study of the epidemiology of bovine anaplasmosis. In: Proceedings of the 6th International Symposium on Veterinary Epidemiology and Economics.

Hungerford, L.L. (1991b). Use of spatial statistics to identify and test significance in geographic disease patterns. Preventive Veterinary Medicine 11 (3-4): 237-42.

Jacqmin-Gadda, H., Commenges, D., Nejjari, C., and Dartigues, J-F. (1997). Tests of geographical correlation with adjustment for explanatory variables: an application to dyspnoea in the elderly. Statistics in Medicine 16: 1283-97.

Jacquez, G.M. (1995). The map comparison problem: tests for the overlap of geographic boundaries. Statistics in Medicine 14: 2343-61.

Kitron, U. (1998). Landscape ecology and epidemiology of vector-borne diseases: tools for spatial analysis. Journal of Medical Entomology 35 (4): 435-45.

Kitron, U., Bouseman, J.K., and Jones, C.J. (1991). Use of the Arc/Info GIS to study the distribution of Lyme disease ticks in an Illinois county. Preventive Veterinary Medicine 11: 243-48.

Kitron, U. and Kazmierczak, J.J. (1997). Spatial analysis of the distribution of Lyme disease in Wisconsin. American Journal of Epidemiology 145 (6): 558-66.

Kulldorff, M. 1998. Statistical methods for spatial epidemiology: tests for randomness. In: M. Löytönen, and A. Gatrell (Eds), GIS and health in Europe. Taylor & Francis, London, England.

Kulldorff, M., Athas, W.F., Feuer, E.J., Miller, B.A., and Key, C.R. (1998). Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos, New Mexico. American Journal of Public Health 88 (9): 1377-80.

Kulldorff, M., Feuer, E.J., and Freedman, L.S. (1997). Breast cancer clusters in the Northest United States: A geographic analysis. American Journal of Epidemiology 146 (2): 161-70.

Kulldorff, M. and Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. Statistics in Medicine 14: 799-810.

Kulldorff, M., Rand, K., and Williams, KG. (1997). SaTScan, version 1.0, software for the spatial and space-time scan statistics. Bethesda, MD: National Cancer Institute [Available at: http://dcp.nci.nih.gov/bb/SaTScan.html].

Langford, I.H. (1994). Using empirical Bayes estimates in the geographical analysis of disease risk. Area 26 (2): 142-49.

Langford, I.H., Bentham, G., and McDonald, A. (1998). Multi-level modelling of geographically aggregated health data: a case study on malignant melanoma motality and UV exposure in the European Community. Statistics in Medicine 17: 41-57.

Laube, P., Stärk, K.D.C., and Keller, H. (1997). In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.13.1-12.13.3.

Laut, P. (1986). Land evaluation for bovine tuberculosis eradication in northern Australia. Australian Geographical Studies 24 (2): 259-271.

Le Sueur, D., Ngxongo, S., Stuttaford, M., Sharp, B., Maharaj, R., Martin, C., and Brown, D. (1995). Towards a rural information system. In: GIS for Health and the Environment. Proceedings of an International Workshop. International Development Research Centre, Ottawa, Canada. 35-51.Levine, N., K.E. Kim, and L.H. Nitz. (1995a). Spatial analysis of

Honolulu motor vehicle crashes: I. Spatial patterns. Accident Analysis & Prevention 27 (5): 663-74.

Levine, N., K.E. Kim, and L.H. Nitz. (1995b). Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accident Analysis & Prevention 27 (5): 675-85.

Mackereth, G. (1998). Spatial data requirements for animal disease management in New Zealand. Unpublished Masters thesis, Massey University, New Zealand.

Maguire, D.J., Goodchild, M.F., and Rhind, D.W.E. (1991). Geographical Information Systems: Principles and Applications. Longman Geoinformation, Cambridge.

Marshall, R.J. (1991). A review of methods for the statistical analysis of spatial patterns of disease. Journal of the Royal Statistical Society 154 (3): 421-41.

McGarigal, K., and Marks, B.J. (1995). FRAGSTATS. Spatial pattern analysis program for quantifying landscape structure Ver. 1.0. General Technical Report PNW-GTR-351. Forest Science Department, Oregon State University, Corvallis, Oregon, U.S.A.

McKenzie, J.S. (1996). The application of GIS to animal health issues using TB management and research in New Zealand as an example. In: J. S. McKenzie (Ed), Proceedings of the Epidemiology and State Veterinary Programmes, NZVA/AVA Pan Pacific Conference. Dept of Veterinary Clinical Sciences, Massey University, Palmerston North, New Zealand. 107-116

McKenzie, J.S., Morris, R.S., Tutty, C.J., and Pfeiffer, D.U. (1997). EpiMAN-TB, a decision support system using spatial information for the management of tuberculosis in cattle and deer in New Zealand. In: R. T. Pascoe (Ed), Proceedings of the Second Annual Conference on Geocomputation. Spatial Information Research Centre, University of Otago, Dunedin, New Zealand. 271-77.

Mott, K.E., Nuttall, I., Desjeux, P., and Cattand, P. (1995). New geographical approaches to control of some parasitic zoonoses. Bulletin of the World Health Organization 73 (2): 247-57.

O'Dwyer, L. (1998). The use of GIS in identifying risk of lead poisoning in Australia. In: P. Firns (Ed), Proceedings of the 10th Annual Colloquium of the Spatial Information Research Centre. University of Otago, Dunedin. 247-58.

Openshaw, S., Charlton, M., Wymer, C., and Craft, A.W. (1987). A mark 1 analysis machine for the automated analysis of point data sets. International Journal of Geographical Information Systems 1: 335-58.

Oranga, H.M. (1995). A GIS approach to the determination of catchment populations around local health facilities in developing countries. GIS for Health and the Environment. Proceedings of an International Workshop. International Development Research Centre, Ottawa, Canada. 52-57.

Paré, J., Carpenter, T.E. and Thurmond, M.C. (1996). Analysis of spatial and temporal clustering of horses with *Salmonella krefeld* in an intensive care unit of a veterinary hospital. Journal of the American Veterinary Medical Association 209 (3): 626-8.

Pereira, J.M.C. and Itami, R.M. (1991). GIS-based habitat modelling using logistic multiple regression: a study of the Mt. Graham red squirrel. Photogrammetric Engineering and Remote Sensing 57 (11): 1475-86.

Pfeiffer, D. U. (1994). The role of a wildlife reservoir in the epidemiology of bovine tuberculosis. Unpublished PhD thesis, Massey University, New Zealand.

Pfeiffer, D. U. (1996). Issues related to handling of spatial data. In: J. S. McKenzie (Ed), Proceedings of the Epidemiology and State Veterinary Programmes, NZVA/AVA Pan Pacific Conference. Dept of Veterinary Clinical Sciences, Massey University, Palmerston North, New Zealand. 83-105.

Pfeiffer, D.U., Duchateau, L., Kruska, R.L., Ushewokunze-Obatolu, U., and Perry, B.D. (1997). A spatially predictive logistic regression model for occurrence of theileriosis outbreaks in Zimbabwe. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.12.1-12.12.3.

Pfeiffer, D.U.and Morris, R.S. (1994). Spatial analysis techniques in veterinary epidemiology. In: Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics.

Pirie, R. (1998). Giardiasis in Auckland water distribution zones. In: P. Firns (Ed), Proceedings of the 10th Annual Colloquium of the Spatial Information Research Centre. University of Otago, Dunedin. 265-70.

Reader, S. (1995a). GIS, health and epidemiology: an annotated resource guide. In: D. de Savigny, and P. Wijeyaratne (Eds), GIS for Health and the Environment. International Development Research Centre, Ottawa, Canada. 145-60.

Reader, S. (1995b). The present state of GIS and future trends. In: D. de Savigny, and P. Wijeyaratne (Eds), GIS for Health and the Environment. Proceedings of an International Workshop . International Development Research Centre, Ottawa, Canada. 3-12.

Rodríguez-Lainz, A., Hird, D.W., Carpenter, T.E., and Read, D.H. (1996). Case-control study of papillomatous digital dermatitis in southern California dairy farms. Preventive Veterinary Medicine 28 (2): 117-31.

Rushton, G. and Lolonis, P. (1996). Exploratory spatial analysis of birth defect rates in an urban population. Statistics in Medicine 15 (7-9): 717-26.

Sanson, R.L.and Pearson, A.P. (1997). Agribase - A national spatial farm database. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.16.1 - 12.16.3.

Sanson, R. L., D. U. Pfeiffer, and R. S. Morris. (1991). Geographic information systems: Their application in animal disease control. Revue Scientifique et Technique. Office International des Epizooties. Vol. 10: 179-95.

Sanson, R.L., Stern, M.W., and Morris, R.S. (1994). Interspread - a spatial stochastic simulation model of epidemic foot-and-mouth disease. In: Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics. 493-95.

Scott, H.M., Ellahoj, E., Wilson, D.J., Martin, S.W., Soskolne, C.L., Coppock, R.W., Guidotti, T.L., Hrudey, S.E., and Lissemore, K.D. (1998). Geographical information systems for integrating air pollution modelling with outcomes and confounders: effects of sour gas processing plant emissions on the health and productivity of cattle. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.14.1-12.14.3.

Smans, M., and Estève, J. (1993). Practical approaches to disease mapping. In: P. Elliott, J. Cuzick, D. English, and R. Stern (Eds), Geographical and environmental epidemiology: methods for small-area studies. Oxford University Press, Oxford, England. 141-150.

Smith, T., Charlwood, J.D., Takken, W., Tanner, M., and Spiegelhalter, D.J. (1995). Mapping the densities of malaria vectors within a single village. Acta Tropica 59: 1-18.

Thomas, R. W. (1990). Introduction: Issues in spatial epidemiology. In: R. W. Thomas (Ed), Spatial epidemiology. Pion Limited, London. 1-14.

Thomas, R.W.E. (1990). Spatial Epidemiology. Pion Limited, London.

Thomson, M.C., Connor, S.J., Alessandro, U.D., Rowlingson, B., Diggle, P., Cresswell, M., and Greenwood, B.M. (1998). A spatial model of malaria risk in The Gambia: predicting the impact of insecticide treated and untreated bednets on malaria infection. In: P. Firns

(Ed), Proceedings of the 10th Annual Colloquium of the Spatial Information Research Centre. Spatial Information Research Centre, Dunedin, New Zealand. 321-28.

Thomson, M., Connor, S., Bennett, S., D'Alessandro, U., Milligan, P., Aikins, M., Langerock, P., Jawara, M., and Greenwood, B. (1996). Geographical persepctives on bednet use and malaria transmission in the Gambia, West Africa. Social Science & Medicine 43 (1): 101-12.

Tinline, R. R. (1988). Persistence of rabies in wildlife. In: J. B. Campbell and K. M. Charlton (Eds), Rabies. Kluwer Academic Publishers, Boston. 301-22.

Tjøstheim, D. (1978). A measure of association for spatial variables. Biometrika 65 (1):109-14.

Twigg, L. (1990). Health based geographical information systems: their potential in the light of existing data sources. Social Science & Medicine 30 (1): 143-55.

Verhasselt, Y. (1993). Geography of health: some trends and persepctives. Social Science & Medicine 36 (2): 119-23.

Walter, S.D. (1993a). Assessing spatial patterns in disease rates. Statistics in Medicine 12 (19-20): 1885-94.

Walter, S.D. (1993b). Visual and statistical assessment of spatial clustering in mapped data. Statistics in Medicine 12 (19-20): 1275-91.

Walter, S.D. and Birnie, S.E. (1991). Mapping mortality and morbidity patterns: an international comparison. International Journal of Epidemiology 20 (3): 678-89.

Ward, M.P., Carpenter, T.E. and Johnson, S.J. (1996). Spatial analysis of seroconversion of sentinel cattle to bluetongue viruses in Queensland. Australian Veterinary Journal 74 (2): 128-31.

Williams, B., D. Rogers, G. Staton, B. Ripley, and T. Booth. (1994). Statistical modelling of geo-referenced data: Mapping tsetse distributions in Zimbabwe using climate and vegetation data. In: B. D. Perry, and J. W. Hansen (Eds), Modelling vector-borne and other parasitic diseases. The International Laboratory for Research on Animal Diseases, Nairobi, Kenya. 267-80.

Wilson, M.L. (1998). Distribution and abundance of *Ixodes scapularis* (Acari: Ixodidae) in North America: ecological processes and spatial analysis. Journal of Medical Entomology 35 (4) 4: 446-57.

CHAPTER 2

Statistical modelling of habitat associated with tuberculous possums in New Zealand

Abstract

This paper describes a case control study of 132 possum capture locations to identify habitat and topographic factors associated with the presence of single or multiple tuberculous possums within a 50-metre square area surrounding each capture location. Data were collected on plant species and structural features of vegetation, the abundance and quality of available den sites, and topographic factors such as slope and aspect. Two unconditional logistic regression models that included different sets of den variables were produced to predict the probability that a location was a multiple TB site rather than a negative TB site. One logistic regression model was produced to predict the probability that a location was a single TB site rather than a negative TB site. Multiple tuberculous possums were more likely to be captured on flatter land with multiple fully enclosed dens or more abundant open dens in trees. Single tuberculous possums were more likely to be captured on flatter land with taller trees, a lower percent coverage of vegetation in tier 5 (height 0.3 – 2 metres), and a higher percentage of ground-level vegetation. The results of this study suggest that clustering of tuberculous possums is associated with a higher density of favourable possum dens, by means of their influence on the probability of contact around the denning area.

Introduction

The brushtail possum (*Trichosurus vulpecula*) acts as a reservoir *of Mycobacterium bovis* infection for farmed livestock in New Zealand (Anon, 1986; Morris et al., 1994; Jackson, 1995). Areas in which tuberculosis (TB) is endemic in possum populations, known as vector risk areas (VRAs), were estimated to cover about 23 per cent of New Zealand in 1995 (O'Neil and Pharo, 1995). This reservoir of infection is severely hampering efforts to eradicate TB from farmed animals in New Zealand and has resulted in a very expensive TB management programme due to the additional cost of controlling possum populations. A total of $26 million was spent on possum control in the 1996/97 financial year and this is expected to increase to almost $30 million by the year 2000 (O'Neil and Pharo, 1995). Current possum control strategies have successfully reduced the incidence of TB in cattle and deer and have reduced the proportion of farms under quarantine for TB control purposes (Pannett, 1995; Livingstone, 1997). However, they have not yet eradicated TB from possum populations in many areas and continual control of possum populations is necessary to keep the incidence of TB in deer and cattle at a low level. As the incidence of TB reactors within cattle and deer

populations in VRAs continues to decline due to the increasing area of land brought under possum control, the marginal return on dollars invested in existing possum control measures will decrease, and it may become increasingly difficult to motivate stakeholders to maintain the current level of funding into the future. It is therefore important to explore more cost-effective ways of maintaining a low incidence and/or of eradicating TB from possum populations. We believe that having information on the habitat in which tuberculous possums are more likely to be located will improve the cost-effectiveness of long term possum control programmes, by facilitating differential targeting of resources at various habitats based on assessment of TB risk.

When considered across a broad geographical scale the estimated point prevalence of grossly recognisable TB amongst possums in endemic areas where populations have not been controlled has commonly been found to be quite low. Hickling (1991, 1995) reported results of 1-2% from a number of surveys. However, there have been a small number of large-scale surveys in which the prevalence has been of the order of 6-8% (Hickling, 1995). At a smaller geographical scale, TB in possums is highly clustered both in space (Hickling, 1995) and in time (Pfeiffer, 1994). In a study that combined data from five large cross-sectional surveys, Hickling (1995) estimated the prevalence of TB within infected subpopulations to range from 9 to 32% (the size of these populations was not specified). During the first four years of a longitudinal study of an infected possum population at Castlepoint, on the south-east coast of the North Island, Pfeiffer (1994) and Jackson (1995) found that the monthly prevalence of TB amongst the 150 study possums varied from 0 to 20%. The prevalence of possums found in annual cross-sectional surveys conducted over a period of 3 days on the forest margin at Ahaura on the west coast of the South Island varied from 4% to an exceptional 53% (Coleman et al., 1994; Coleman et al., 1996).

The most detailed scale of clustering has been reported at the level of individual traps (Hickling, 1995) and individual dens (Pfeiffer, 1994). Hickling (1995) used a crowding index as a measure of 'patchiness' of TB possums between individual traps in a line and showed that tuberculous possums were 4-16 times as crowded as if they were distributed randomly. Pfeiffer (1994) showed that the median distance between dens used by tuberculous possums was 25 metres, compared to a median distance of 55 metres between dens used by non-tuberculous possums and the nearest den used by a tuberculous possum. Pfeiffer et al. (1995) reported on a survey that was conducted using trap lines on forest-pasture boundaries in the central North Island, in which the average size of a cluster was 33 ± 25 metres. One cluster

covered an area with a diameter of 120 metres. Hickling (1991) reported on two cross-sectional surveys of possums in the Hohonu area (located on the west coast of the South Island) which were conducted 17 years apart. While there were fewer TB possums caught during the second survey, the TB possums were caught either at the same trap or one trap either side of traps (set 20 metres apart) which had caught TB possums during the first survey, giving cluster sizes that varied from 20 to 40 metres. In all the calculations of cluster size from cross-sectional surveys, it was not known whether the trap lines crossed a particular cluster right through the middle or crossed at less than the full diameter. Information from these studies plus a small number of other studies, indicate that the size of possum TB clusters is relatively small, with a cross-sectional width of the order of 20-40 metres.

Data from three of the four longitudinal or repeated cross-sectional studies of infected possum populations conducted in New Zealand suggest that some clusters are spatially very stable over time, albeit with a fluctuating prevalence, while others are more sporadic in nature, existing for a variable period then disappearing. Some possum TB clusters in the two Hohonu surveys were stable over 17 years, while others disappeared in the intervening period (Hickling, 1991). Furthermore, the stable clusters remained in the same small focus, and there was no evidence of outward spread from these. The main cluster of TB possums at the Castlepoint study site remained in the same 2 to 4 hectare area throughout the 9-year course of the project (Pfeiffer, 1994; Jackson, 1995; Lugton, 1997; Corner, pers comm 1999). The study at Hohotaka, in the central North Island involved observations on an infected possum population that was subject to annual control (Hickling, 1989; Caley, 1995a). Despite the population being maintained at 20-40% of its pre-control size, tuberculous possums were caught at the original two cluster sites during 4 of the 6 years following initial control. The infected possum population at Ahaura was found to be more homogeneous in its distribution during the survey conducted in 1992 when 54% of 68 possums were tuberculous. However, in subsequent years there was evidence of persistent clustering across 2 or 3 years at two trap locations, whilst TB was more transient at other trap locations where tuberculous possums had been caught, with no persistence across years (Coleman et al., 1994; Coleman et al., 1996).

Information on the spatial and temporal patterns of different restriction endonuclease analysis (REA) types of *M. bovis* found in the Castlepoint study provides further insights into the nature of possum TB clusters. The spatial and temporal distribution of four REA types found in the first 5 years of the study (Pfeiffer, 1994; Jackson, 1995) and the temporal distribution of

8 different REA types found during the first 5½ years of the longitudinal study (Lugton, 1997) show a variation in the patterns of different REA types. Possums with three different REA types denned within the location of the main TB cluster during the first four years of the study, but the temporal patterns of prevalence of each REA type were varied. Lugton (1997) suggested that "the most likely explanation for this phenomenon was natural variation in the opportunities for transmission presented to particular infected individuals." One less common REA type appeared as a temporary cluster in another part of the study area. This cluster involved two juvenile possums that appeared to have been infected in an area outside the study site before immigrating, and died on the study site without spreading infection to other study possums. Langmuir (1961, cited by Morris et al., 1994) notes that "Epidemiologically, the dominant concept in tuberculosis control has been, and to a large degree still is, the importance of close contact." He goes on to say that "These observations are much more consistent with the hypothesis that only certain tuberculous individuals act as effective disseminators and these do so probably intermittently and only under certain circumstances." The spatial and temporal patterns of TB in possums suggest an interaction between consistent spatial factors that influence where tuberculous possums are more likely to be located, perhaps by influencing the likelihood of contact and effective disease transmission, and less consistent temporal factors that influence the infectivity of tuberculous possums.

These data support the hypothesis proposed by McKenzie and Morris (1995) that persistent clusters occur at locations where TB spreads well between possums and thus becomes endemically established in the local possum population. Although the prevalence of possums with detectable lesions of TB sometimes falls to a very low level at these locations, there is a sufficiently high prevalence of "subclinically" infected possums to maintain the disease at the location. Temporary or sporadic TB clusters occur where there is much less spread of TB amongst possums, and the disease does not become endemically established in the local possum population. Hickling (1995) found that the proportion of habitat infested with tuberculous possums was highly correlated with mean crowding. He suggests that "when possum populations are crowded, tuberculosis spreads out from foci of infection to the surrounding population. Conversely, when possum numbers are controlled the disease may retreat into small foci where the disease remains locally prevalent." This is a similar pattern to the spatial and temporal patterns demonstrated by the two major REA types of *M. bovis* found at Castlepoint. These data suggest that persistent cluster locations are critical to the long term maintenance of TB within possum populations, and are the most likely locations for TB to

persist following the control of possum populations.

Pfeiffer (1994) and Jackson (1995) found that TB clustering was associated with possum denning areas. This provided strong evidence that transmission of TB amongst possums was occurring as a result of social interactions in the vicinity of dens, rather than in other activity areas such as those used for feeding. An association between possum TB clusters and habitat was found in the possum surveys conducted in the Hohonu. Hickling (1991) suggested that the high TB prevalence in the first 200 metres of forest at both surveys may be explained by habitat factors favouring disease transmission at certain sites, such as heavily-used dens or areas of palatable pasture vegetation where possums congregate to feed. Cook's (1975) report on the first cross-sectional survey of possums conducted in the Hohonu, noted that transmission of TB was most likely to occur in the band of podocarp forest located 50 – 400 metres from the lowest margin of the forest. Some authors have reported an association between TB possums and the bush-pasture margin (Davidson, 1976; Hickling, 1989). In his paper on the 1973 survey conducted in the Hohonu, Coleman (1988) suggested an association between TB and the bush pasture margin which is where he observed the highest prevalence of TB. However, in his report of the same survey, Cook (1975) wrote that "the prevalence of tuberculosis infection was highest (15%) on pasture lands but was reduced to 8.4% at the forest-pasture margin. (As a convention for the purpose of this trial, the forest/pasture margin was considered to extend 50 m into the forest). The rate rose to 12.5% between 51 and 400 m into the forest and then fell to 11.3% at between 401m and 2000m and to 4% at between 2001 and 3600m." An association between the forest/pasture margin and TB prevalence in possums has not been found in other studies. Coleman et al. (1992) found the prevalence of TB in possums caught on 14 lines running 1500 metres into the forest from the forest/scrub margin in Ahaura did not decrease with distance from the forest edge. In a study to identify the effectiveness of different buffer widths for possum control operations in the central North Island, a TB prevalence of 2.2% of 90 possums was found before control in the deep forest section of the 7-km buffer compared with zero at the forest/pasture margin (Coleman et al., 1998). Subsequently, 2.2% of 89 possums was found with TB lesions in the deep forest section of the 3-km buffer two years after the implementation of possum control, compared with zero at the forest/pasture margin.

Pfeiffer (1994) found that distance to the nearest den used by a tuberculous possum, height above sea level, aspect of the den site and interaction between aspect and height all influenced the likelihood that a particular den site would have been used by a tuberculous possum. A

high proportion of dens used by tuberculous possums was located in the northern part of the study site, and Pfeiffer postulated that the expression of TB in possums was influenced by the harshness of the site due to environmental exposure. However, Lugton (1997) found that body weight was higher and serum cortisol levels were lower in possums in this northern part of the study area compared with the two areas where a lower density of possum dens was found. These results suggest that possums in this apparently harsh area of the study site were less stressed than in other areas, and that the disease presence was associated with environmental factors other than those which might induce stress in possums. It is possible that the exposure of the steep slope to strong winds resulted in habitat cover that provided favoured denning areas for possums, leading to a higher density of possums in this area.

Caley (1996) conducted a study that attempted to address the hypothesis that possums with clinical TB were more likely to be found in areas with poorer quality dens which created a more stressful environment. He compared the quality and location of dens used by possums in an area where tuberculous possums had previously been caught and an area where no tuberculous possums had been caught over a seven-year period. He found no significant difference in the quality of dens between the two sites. However, there was a higher proportion of dens located above the ground in the non-TB area (58% versus 14%). During the habitat study reported in this paper we collected data from the non-treatment area included in Caley's study, and it had the characteristics of a multiple TB site with many fully enclosed dens available and low slope. However TB has not been found at this site in the course of the seven-year study conducted at Hohotaka, despite TB being endemic in possums for a minimum of five years at a site 2 kilometres away. A possible explanation for TB not becoming established at the negative site studied by Caley (1996) is that the higher proportion of den sites above ground did not provide an environment conducive to the transmission of TB. Likewise, this factor may explain why TB did not become established at the Orongorongo study site at the southern end of the North Island (Brockie et al., 1987). In a two year study at this location, Cowan (1989) found only 2% of 1987 denning events were located at or below ground level and den sharing was not common. On only 3.4% of 1987 occasions were 2 or more possums located simultaneously in the same den tree, and on at least 70% of these occasions radio signal characteristics suggested that possums were in separate dens in the same tree. Most den trees were large and had several clumps of perching epiphytes.

There is no evidence of a positive relationship between the prevalence of TB in possums and the density of the possum population at the scale of large cross-sectional surveys (Hickling,

1995), or at the smaller scale of 1-km^2 sampling units within such surveys (Hickling, 1989). At Ahaura, the prevalence of TB dropped over the 5 years that it was studied despite a steady increase in the abundance of the trapped population each year, providing evidence that TB infection can remain and rebuild even in very low-density populations (Coleman et al., 1994; Coleman et al., 1996). Pfeiffer (1994) showed that the area of the highest density of traps that caught the highest proportion of tuberculous possums in the total catch per trap, was different to the area with the highest density of total possum catch. There is evidence of a density effect at a very localised scale. Hickling (1995) found that disease prevalence was 2.2 times as high at trap sites where 3 or 4 possums were collected as at trap sites with only 1 or 2 possums. These data show that geographical scale is very important in relation to identifying an association between possum density and clustering of TB. The association is strongest at a very localised scale, which appears to be of the order of less than one hectare.

Multivariate statistical modelling has been used to explore the relationship between the distribution of wildlife species and specific habitat features, and to predict the distribution of species (Broschart et al., 1989; Lindenmayer et al., 1990; Pereira and Itami, 1991; Buckland and Elston, 1993), and of subpopulations within species (Bian and West, 1997). In this paper we describe the application of logistic regression modelling in a study to identify habitat factors that could be used to predict the locations of tuberculous possums, and to gain further insights into the epidemiology of TB in possums.

Materials and Methods

Study population

We used a case control study design in which the unit of interest was a 50 metre square area surrounding a site at which a possum of known TB status had been captured in an earlier investigation. Study sites were selected from among a number of cross-sectional possum TB surveys that had been conducted by ourselves, Ministry of Agriculture and Fisheries' staff and private trappers. We used surveys for which the locations of TB-positive possums and of the trap or poison lines had either been recorded or could be recalled to an accuracy of 20-30 metres. A "TB-positive" site was defined as one where one or more of the possums caught at the site had gross TB lesions on post mortem examination. TB-positive sites were further classified as "single TB-positive" and "multiple TB-positive" sites. A single TB site was one where only one TB-positive possum had been caught within a 50-metre radius of the site. A

multiple TB site was where more than one TB-positive possum had been caught. A "TB-negative" site was one where equivalent intensity of investigation had been conducted but no TB-positive possums had been caught in the immediate vicinity at any time. Negative sites were a minimum of 100 metres and a maximum of 1500 metres from the nearest boundary of a TB-positive site. One TB-negative site was selected for every positive site from within the same survey area, generally on the same farm or a neighbouring farm. This method was adopted to increase the chance that possums at negative sites had the same putative risk of being exposed to TB as those at positive sites. Negative sites were randomly selected from all eligible sites within the survey area. A total of 132 sites was included in the study, comprising 66 TB-positive sites and 66 TB-negative sites. The TB-positive sites comprised 25 multiple and 41 single TB sites.

Independent variables

We collected data on three groupings of environmental factors: vegetation, den and topographic factors. Vegetation data was collected using the Recce inventory method (Allen, 1992) which was designed by the New Zealand Forest Research Institute for rapid broad-scale surveys of compositional variation in mountain forests. Vegetation factors were the abundance of each plant species within five different height tiers (<0.3, 0.3-2, 3-5, 6-12, >12 metres), mean height of the tallest trees or shrubs at the site, abundance of epiphytes, and presence or absence of large-diameter trees (trees with a circumference at breast height (cbh) greater than 75cm). Trees with a cbh greater than 75cm were considered large enough to contain potential possum den sites (Green and Coleman, 1987). The abundance of plant species within each tier was calculated by estimating the percent of the site covered by each species if they were projected down to ground level, and scoring this using the Braun-Blanquet scale with six categories (Kent & Coker, 1992). The score for each cover class plus the range and midpoint of the proportion of cover represented by each category is listed in Table 1. The proportion of the ground covered by vascular vegetation, moss, water, large logs and roots, and bare ground was also recorded. The abundance of epiphytes was recorded on a subjective scale of 1 (rare) – 5 (abundant), following the ACFOR scale described by Kent and Coker (1992).

Table 1. Details of the Braun-Blanquet scale used to score the abundance of plant species within each height tier.

Score	Percent coverage of site	Midpoint of abundance class
1	<1	0.5
2	1-5	3
3	6-25	15.5
4	26-50	38
5	51-75	63
6	76-100	88

The plant species data at each site was summarised in three ways. Firstly, the mid-point of the abundance class for each species was summed to give a total abundance score for vegetation cover within each height tier. Secondly, sites were classified according to the predominant group of plant species present at each site using a cluster analysis which was conducted using PC-Recce, a vegetation inventory data analysis package produced by the Forest Research Institute (Hall, 1992). This is a package of computer programmes designed to check, summarise and analyse vegetation description data recorded using Recce inventory methods. PC-Recce includes programmes for clustering, ordinating and classifying data. A Quantified Sorensen's K coefficient was used to measure similarity between sites, and the group average was used as the linkage method to combine sites within a cluster. The weighting assigned to each species in each tier at each site was set equal to the score for the cover class recorded for each species, which ranged from 1 to 6 (see Table 1). One cluster analysis included vegetation data from tiers 2 to 5, which represented plant species in the canopy, sub-canopy and understorey layers of each site. A second cluster analysis included vegetation data from only tiers 2 and 3, representing the canopy and sub-canopy layers of each site. The ground tier (height <30cm) was not included in either analysis as we had not recorded individual species at the ground level. Neither cluster analysis produced a set of plant species' groupings that could be used to classify sites for subsequent multivariate statistical analyses, as the composition of plant species at the study sites was too diverse to create a small number of species categories that each contained a sufficient number of sites.

Thirdly, plant species within tiers 2 and 3 were classified into two groups on the basis of whether the species was likely to provide fully enclosed dens, either within their trunks or under their roots. These were classified either as 'good den trees' or 'poor den trees'. A

description of the den tree classification of each species is presented in Table 2. The total abundance of each category was calculated by summing the mid-point of the abundance class for each species within each category.

Table 2. List of tree species located in tiers 2 and 3 that were classified as either 'good den trees' or 'poor den trees'.

Good den trees	Poor den trees
Rimu (Dacrydium cuppressinum)	Titoki (Alectryon excelsus)
Totara (Podocarpus totara)	Beech (Nothofagus spp)
Miro (Prumnopitys ferruginea)	Eucalypt
Kahikatea (Podocarpus dacrydioides)	Rewarewa (Knightsia excelsa)
Hinau (Elaeocarpus dentarus)	Kanuka (Kunzea ericoides)
Mahoe (Melicytus ramiflorus)	Manuka (Leptospermum scoparium)
Willow (Salix species)	Red matipo (Myrsine australis)
	Toro (Myrsine salicina)
	Black maire (Nestegis cunninghamii)
	Tanekaha (Phyllocladus trichomanoides)
	Pine (Pinus radiata)
	Pittosporum species
	Poplar (Populus spp)
	Lancewood (Pseudopanax crassifolius)
	Kowhai (Sophora spp)
	Wineberry (Aristotelia serrata)
	Tawa (Beilschmiedia tawa)
	Rangiora (Brachyglottis repanda)
	Putaputaweta (Carpodetus serratus)
	Coprosma species
	Pigeonwood (Hedyarya arborea)
	Tree fuchsia (Fuchsia excorticata)
	Lacebark (Hoheria populnea)
	Ngaio (Myoporum laetum)
	Heketara (*Olearia rani*)
	Kaikomako (*Pennantia corymbosa*)
	Five finger (*Pseudopanax arboreus*)

Data on the location, abundance and quality of available possum dens was collected at each site by searching for available dens for up to 30 minutes, using a torch to look into dark areas. The quality of a den site was scored using the same four point scale as that used by Cole (1993) and Caley et al. (1998). Scores ranged from 1 to 4 depending on the degree to which the den was exposed to the environmental elements, with one being completely exposed and 4 being completely enclosed (Table 3).

Table 3. Description of the four categories used to score the quality of dens.

Score	Description
1	Very open to the environment, direct sunlight, rain and wind, with no structure defining a refuge site.
2	Open to direct sunlight, rain and wind, with some degree of protection offered by a 'structure' e.g. hole in the side of a trunk in which possums can sit.
3	Mostly protected but still allowing daylight to enter. Protected from the rain, wind and sunlight in all but extreme conditions.
4	Totally protected from the elements and represented an almost entirely enclosed air space.

The abundance of dens of quality 2, 3 and 4 was scored and recorded for each of 21 locations at every study site. Abundance was scored using a subjective scale of 1 (rare) – 5 (abundant), according to the ACFOR scale described by Kent and Coker (1992). Den locations were: under live tree roots, under dead tree roots, in live tree trunks, in dead tree trunks or stumps, in hollow logs, under logs, under piles of fallen tree fern leaves, under piles of dead branches and leaf litter, in recesses within banks, in recesses within rocks, in rabbit burrows, under gorse (*Ulex europaeus*), under flax (*Phormium tenax*), under blackberry (*Rubus fruticosis*), under toetoe (*Cortaderia toetoe*), under *Carex secta*, under raupo (*Typha muelleri*), under dense ground fern, under long grass, under bracken (*Pteridium esculentum*), and other. The abundance of quality 1 dens was scored at the following locations: in the crowns of tree ferns, in epiphytes, in large trees that had flat areas at the top of the trunk or in the forks between large branches and the trunk or flat areas on large branches where possums could sit comfortably. These most commonly included Hall's totara (*Podocarpus cunninghamii*), totara (*Podocarpus totara*), rimu (*Dacrydium cupressinum*), willow (*Salix* species), macrocarpa (*Cupressus macrocarpa*) and cabbage trees (*Cordyline australis*). Quality one dens were not coded at the different locations in which they were found as these were all considered to be 'trees'. The den data at each site was summarised for the analysis by summing the abundance score for dens of the same quality across all locations, giving a total abundance score of dens

of each quality at each site. Den data was represented in a second way as a binary variable based on the presence or absence of multiple dens of each quality.

Topographic variables were the mean slope and aspect of each site. Slope was measured using a clinometer laid on a one-metre ruler that was laid on a representative area of slope. Where the slope varied within the site, an average value was estimated for the whole site. Aspect was measured using a compass and readings were adjusted for a 21° difference between magnetic and grid north. Aspect was divided into five categories comprising flat (no aspect), north (316-45°), east (46-135°), south (136-225°), west (226-315°), and flat.
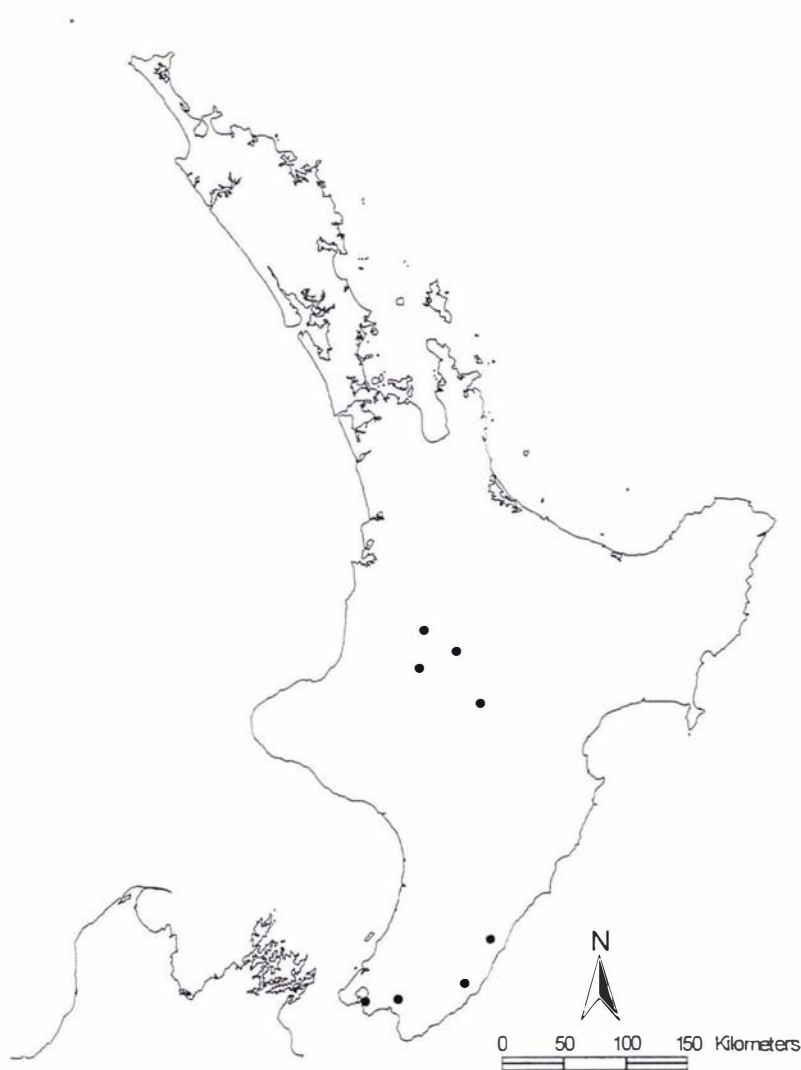
Figure 1. Map of the North Island of New Zealand showing general locations of the study sites.

Statistical analysis

Univariate analyses were used to screen the relationship between individual variables and TB status of the sites. Chi-squared analysis was used for categorical variables and a Mann-Whitney U test for difference between medians was used for continuous variables as none of these was normally distributed. Variables that were associated with TB status of the sites at p < 0.2 were used in multivariate analyses.

Table 4: Results of univariate chi-squared analyses for categorical variables.

Variable	Multiple (n=25)		Single (n=41)		Negative (n=66)		Total	p-value
	Obs	Exp	Obs	Exp	Obs	Exp		
Epiphytes								
Yes	6	4	8	6	7	10	21	
No	19	21	33	35	59	56	111	0.27
Large-diameter trees								
Yes	17	15	26	24	34	39	77	
No	8	10	15	17	32	27	55	0.19
Aspect								
Flat	8	5	12	8	6	13	26	
North	5	5	7	8	14	13	26	
East	4	4	7	7	12	11	23	
South	6	6	6	9	17	14	29	
West	2	5	9	9	17	14	28	0.16
Multiple quality1 dens								
Yes	18	13	19	21	31	34	68	
No	7	12	22	20	35	32	64	0.09
Multiple quality2 dens								
Yes	21	23	35	37	64	60	120	
No	4	2	6	4	2	6	12	0.06
Multiple quality3 dens								
Yes	20	18	31	30	46	48	97	
No	5	7	10	11	20	18	35	0.66
Multiple quality4 dens								
Yes	19	12	15	19	27	31	61	
No	6	13	26	22	39	36	71	0.005

Obs: Observed   Exp: Expected

Separate unconditional logistic regression models were produced for "multiple TB-positive" sites and for "single TB-positive" sites, using the habitat factors that were selected from the univariate analyses. To overcome multicollinearity between the two groups of den variables, two separate groups of habitat factors were modelled for each of the "multiple TB-positive" and "single TB-positive" sites. One group included dens represented as the total abundance of dens of each quality score plus vegetation and topographic variables, while the other included dens represented as the presence or absence of multiple dens of each score plus vegetation and

Table 5. Summary statistics and results of univariate analyses using the Mann-Whitney U test for all continuous variables for which data was collected in the habitat study.

Variable	Multiple (M) n = 25	Single (S) n = 41	Negative (N) n = 66	Mann-Whitney U results where p < 0.2
Slope				
Median (Range)	10 (0-45)	15 (0-50)	29 (0-50)	M<S, M<N*, S<N*
Std deviation	11	15	13	
Top height				
Median (Range)	18 (7-30)	13 (4-30)	13 (3-35)	M>N*, M>S, S>N
Std deviation	8	8	8	
Percent cover tier 2 (>12m)				
Median (Range[1])	3 (0-108)	3 (0-96)	0.5 (0-120)	M>N
Std deviation	34	26	29	
Percent cover tier 3 (5-12m)				
Median (Range[1])	19 (3-110)	16 (0-103)	22 (0-111)	M>S, N>S
Std deviation	33	29	32	
Percent cover tier 4 (2-5m)				
Median (Range[1])	31 (0-104)	10 (0-99)	31 (0-150)	M>S*, N>S*
Std deviation	33	26	33	
Percent cover tier 5 (.3-2m)				
Median (Range[1])	21 (.5-153)	11 (0-85)	17 (0-114)	M>N, M>S*, N>S*
Std deviation	42	18	32	
Percent cover tier 6 (<0.3m)				
Median (Range[1])	4 (.5-140)	31 (.5-105)	5 (0-99)	M<S*, S>N*
Std deviation	43	36	25	
Percent cover vegetation				
Median (Range)	20 (1-95)	68 (.5-100)	26 (0-100)	M<S*, S>N*
Std deviation	35	34	33	
Percent cover bare ground				
Median (Range)	70 (0-98)	29(0-99)	65 (0-99)	M>S*, S<N*
Std deviation	34	32	34	
Percent cover logs				
Median (Range)	1 (0-48)	0 (0-15)	0.5 (0-15)	M>N*, M>S*
Std deviation	10	4	3	
Percent cover water				
Median (Range)	0 (0-25)	0 (0-5)	0 (0-88)	
Std deviation	5	1	12	
Percent cover moss				
Median (Range)	0 (0-10)	0 (0-56)	0 (0-47)	M<N, N>S
Std deviation	2	12	7	

[1] Values could sum to greater than 100% because of overlap of species within tiers.

* Significant at p α 0.05

Variable	Multiple (M) n = 25	Single (S) n = 41	Negative (N) n = 66	Mann-Whitney U results where p < 0.2
Abundance den quality 1				
Median (Range)	5 (0-20)	1 (0-12)	1 (0-8)	M>N, M>S*
Std deviation	4	3	0.3	
Abundance den quality 2				
Median	4 (0-12)	5 (0-19)	5 (0-19)	M<N
Std deviation	4	5	4	
Abundance den quality 3				
Median	5 (0-16)	4 (0-26)	3 (0-17)	M>N*, M>S, S>N
Std deviation	5	4	4	
Abundance den quality 4				
Median (Range)	4 (0-14)	1 (0-12)	1 (0-17)	M>N*, M>S*
Std deviation	3	3	3	
Percent cover good den trees				
Median (Range)	6 (0-89)	0 (0-66)	0 (0-114)	M>S*, M>N*
Std deviation	24	13	18	
Percent cover poor den trees				
Median (Range)	17 (0-114)	17 (0-103)	21 (0-105)	M<N, S<N
Std Deviation	30	30	35	

[1] Values could sum to greater than 100% because of overlap of species within tiers.

* Significant at p α 0.05

topographic variables. A forward selection process was used with a cut-off p-value of 0.05 for entry into and a p-value of < 0.10 for removal from the model. Two-way interaction terms between the variables that were selected in the final main effects model were tested using the chi-squared goodness of fit statistic.

A Receiver Operating Characteristic (ROC) curve was generated for the regression model that had the highest R^2 for "multiple TB-positive" sites and the model for "single TB-positive" sites. These curves display the sensitivity and specificity of the model based on varying the cut-off value for predictions (cut-off value = probability value), and enable an optimal cut-off point to be identified to achieve a desired sensitivity and specificity level for classifying sites based on the predicted probability of the logistic regression model.

Analyses were conducted using NCSS (v 97, Number Cruncher Statistical Systems, Kaysville, Utah).

Results

The areas in which the 132 study sites were located are shown in Figure 1, with 95 sites in the central North Island and 37 in the south-eastern North Island. Summary descriptive statistics and results of the univariate analyses for categorical variables are shown in Table 4, and those for continuous variables are shown in Table 5. The variables that were put forward for multivariate analyses were: slope, mean top tree height, percent cover in tiers 3, 4 and 5, percent ground covered in vascular vegetation, percent ground covered with large logs and roots, presence/absence large-diameter trees, abundance of good den trees, abundance of poor den trees, abundance of dens of each quality score: 1, 2, 3, and 4 and presence/absence of multiple den sites of each quality score. Several of the independent variables turned out to be highly correlated with another biologically associated variable, and hence produced collinearity problems in the analysis. In each case one of the collinear variables was removed from the analysis, and the factor represented by the remaining variable. Percent cover of tier 2 was excluded as it was highly correlated with mean top height ($r=0.8$). Percent cover in tier 6 was excluded as it represented much the same information content as percent cover of vascular vegetation ($r=0.6$). Percent cover of bare ground was excluded as it was highly negatively correlated with percent cover of vascular vegetation ($r=-0.9$). Percent of the site covered by water and moss were not included in the analyses as the values were very small, so there was insufficient variation to give rise to significant differences between sites. Aspect was not included because the main category contributing to the significant difference between TB status groups was the flat category, and this variable was represented by the slope variable.

Table 6. Final logistic regression model for "multiple TB-positive" sites, using the group of variables in which dens were represented as the abundance of dens of each quality.

Variable	Regression coefficient	Standard deviation	Z-value	P-value	O.R.	95% C.I.
Constant	0.92	0.83	1.10	0.272		
Slope (10° units)	-0.87	0.28	-3.19	0.001	0.42	0.25-0.72
Abundance den1	0.31	0.11	2.91	0.004	1.37	1.11-1.69
Abundance den2	-0.24	0.11	-2.29	0.022	0.79	0.64-0.97
R-squared	0.29					

O.R. = Odds ratio   C.I. = Confidence interval

Table 7. Final logistic regression model for "multiple TB-positive" sites, using the group of variables in which dens were represented as the presence/absence of multiple dens.

Variable	Regression coefficient	Standard deviation	Z-value	P-value	O.R.	95% C.I.
Constant	1.45	1.04	1.38	0.166		
Slope (10° units)	-0.84	0.23	-3.57	0.000	0.43	0.27-0.69
Multiple Den4	1.58	0.61	2.57	0.010	4.86	1.46-16.23
Multiple Den2	-2.02	1.02	-1.99	0.047	0.13	0.02-0.97
R-squared	0.26					

O.R. = Odds ratio   C.I. = Confidence interval

Logistic regression analyses resulted in two significant models being produced for "multiple TB-positive" sites and one model for "single TB-positive" sites. The significant variables in one final model for "multiple TB-positive" sites were slope and abundance of quality 1 and quality 2 dens (Table 6), while those in the second model were slope and the presence of multiple quality 2 and quality 4 dens (Table 7). In both models slope and the abundance or presence of multiple quality 2 dens were negatively associated with the odds of a site being "multiple TB-positive", while the abundance of quality 1 dens and the presence of multiple quality 4 dens were both positively associated with the outcome variable. The multiple TB model in which den factors were represented as the total abundance of dens of each quality had a slightly higher R^2 (0.29) than the model in which den factors were represented as the presence/absence of multiple dens of each quality (0.26). The final model for "single TB-positive" sites included slope, mean top height of vegetation, percent coverage of tier 5 (0.3 – 2 metres high) and percent of the ground covered with vascular vegetation (Table 8). Slope and percent coverage of tier 5 were both negatively associated with the odds of a site being "single TB-positive", while mean top height and percent of the ground covered with vascular vegetation were positively associated.

Table 8. Final logistic regression model for "single TB-positive" sites.

Variable	Regression coefficient	Standard deviation	Z-value	P-value	O.R.	95% C.I.
Constant	-1.41	0.83	-1.70	0.09		
Slope (10° units)	-0.47	0.17	-2.75	0.01	0.63	0.45-0.88
Height (10 metre units)	0.62	0.31	1.99	0.05	1.87	1.01-3.46
T5 cover (10% units)	-0.20	0.10	-2.01	0.05	0.82	0.68-1.00
Ground veg[n] (10% units)	0.30	0.08	3.92	0.00	1.34	1.16-1.56
R-squared	0.24					

O.R. = Odds ratio   C.I. = Confidence interval
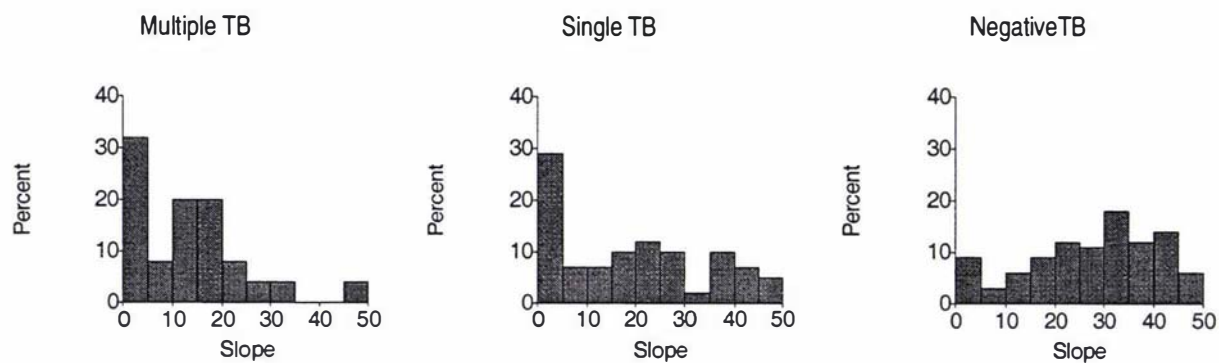
Figure 2. Histograms showing the percent of sites within 5-degree slope categories for each TB status group.
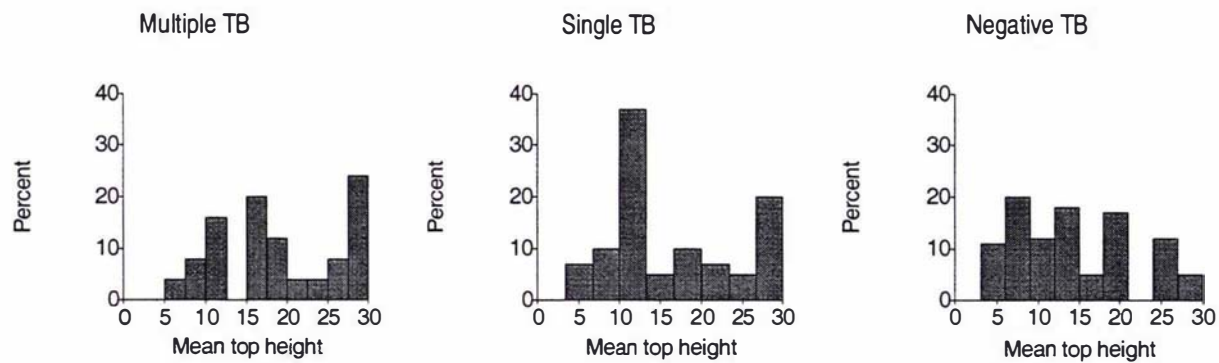
Figure 3. Histograms of the percent sites within 5-metre mean top height categories for each TB status group.
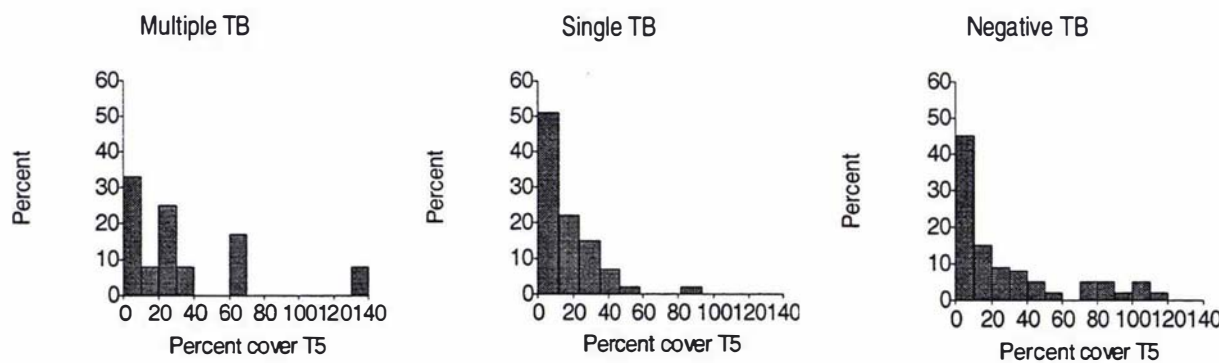
Figure 4. Histograms of the percent sites within 10 percent abundance categories of vegetation cover in tier 5 for each TB status group.

Multiple TB   Single TB   Negative TB
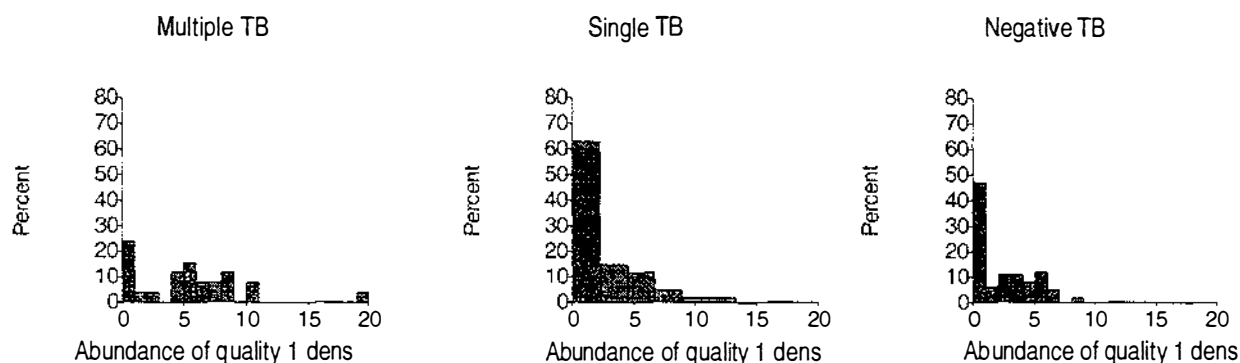
Figure 5 . Histograms showing the percent of sites within categories for the abundance score of quality 1 dens.

Histograms of the percent of study sites within 5-degree slope categories for each of the three TB status groups are shown in Figure 2. Eighty percent of "multiple TB-positive" sites had slopes less than 20° compared with 50% of "single TB-positive" sites and 27% of "TB-negative" sites. Histograms of the percent of study sites within 5-metre height categories are shown in Figure 3. Forty percent of "multiple TB-positive" sites and 35% of "single TB-positive" sites contained trees that were an average of 20 metres or taller, compared with only 15% of "TB-negative" sites. Histograms of the percent of study sites that were within 10% coverage categories for vegetation in tier 5 for each TB status group are shown in Figure 4. Seventeen percent of "TB-negative" sites had more than 60% coverage compared with 2% of "single TB-positive" sites and 16% of "multiple TB-positive" sites. Some values for percent coverage of Tier 5 were greater than 100% due to layering of different species within this height tier. A histogram of the percent of quality 2, 3 and 4 dens that were recorded at each of the 21 locations is shown in Figure 6. This figure shows the higher proportion of quality 2 dens that was located under logs, fallen tree fern leaves, piles of branches or ground fern, or in bank recesses compared with quality 4 dens, of which a higher proportion was located under live or dead tree roots, in holes in live tree trunks or in hollow logs. The distribution of quality 1 dens across different locations is not shown as these were all coded as being located in 'trees'. Histograms of the percent of study sites within abundance categories for quality 1 dens are shown in Figure 5. Forty percent of "multiple TB-positive" sites had an abundance score for quality 1 dens greater than 5 compared with 11% and 10% for "single TB-positive" and "TB-negative" negative sites respectively.
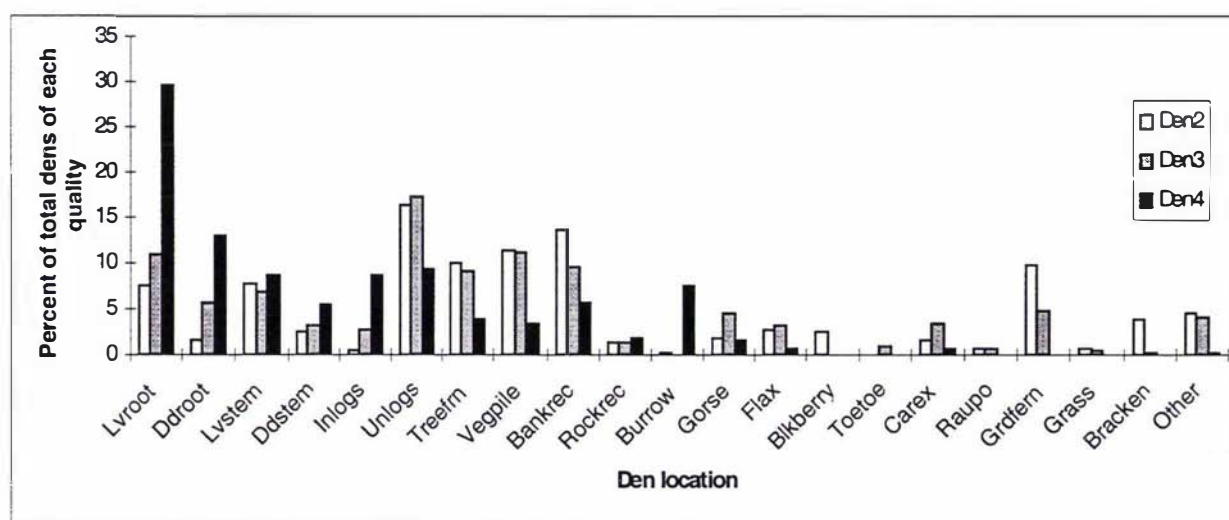
Figure 6. Histogram showing the percent of total dens of each quality at each location.

The ROC curve for the "multiple TB-positive" model that included the abundance of dens and the ROC curve for the "single TB-positive" model are shown in Figure 7. A cut-off probability of 0.26 for the "multiple TB-positive" model and 0.32 for the "single TB-positive" model jointly maximised the sensitivity and specificity of each model. The sensitivity and specificity for the "multiple TB-positive" model using a cut-off probability of 0.26 were 0.92 and 0.74 respectively, and those for the "single TB-positive" model using a cut-off probability of 0.32 were 0.82 and 0.70. These figures show that the "multiple TB-positive" model was able to discriminate multiple TB sites from negative sites more accurately than the "single TB-positive" model was able to discriminate single TB sites from negative sites.

Discussion

Our initial intention had been to identify TB-positive and TB-negative sites for this study by conducting a series of cross-sectional surveys ourselves, so that we had full control over the quality of possum-location data. Because of the difficulties associated with finding tuberculous possums by means of cross-sectional post mortem surveys (Jackson, 1995; Lugton, 1997), we found in conducting a series of such examinations that we were unable to locate a sufficient number of TB-positive sites with the number of investigation hours available. As an alternative, we then changed strategy and made extensive use of existing data from surveys conducted in the past, for which the location of trap or poison lines and the location of TB-positive possums had either been recorded or could be recalled precisely by personnel who conducted the surveys. As a result we had no data on the number of possums

ROC curve of multiple versus negative TB sites

ROC curve of single versus negative TB sites
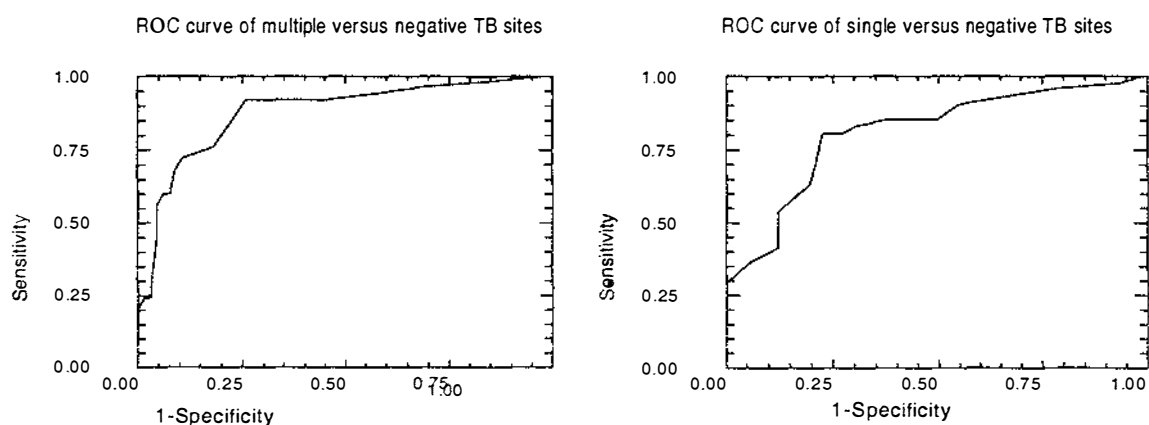
Figure 7: ROC curves for the "multiple TB-positive" logistic regression model that included the abundance of dens and the "single TB-positive" model, showing the sensitivity and specificity with which each model predicts the TB status of a site using different probability values as cut-off points.

caught at each location for most of the surveys used in this study. We thus had no information on the distribution of the population at risk, and the TB-negative sites selected in this study may in some cases have represented sites where no possums were caught. As we only had data on possum capture sites with no denning data for captured possums, we used the capture site to represent the denning area of each possum. We made the assumption that the habitat within 50 metres of the capture location represented that in which the possum normally denned. The size of our area of interest was a compromise between being large enough to capture the possum's most likely denning area, but not so large that the habitat at each site was so varied that different patterns of habitat could not be discerned between sites. We are aware that the data used in this study have inherent weaknesses, and we consider the results in the light of these weaknesses. However, we believe that the number of sites included in this study was large enough to accommodate many of the possible biases in the data, and has enabled us to draw meaningful conclusions from the results.

The multiple logistic regression model provides two sources of information. It identifies the variables that are significantly associated with the outcome variable and estimates the magnitude of their effects, expressed by the odds ratio, controlling for the other variables in the model. This provides insights into the epidemiology of the disease. It also provides a statistical model that can be used to predict the probability of particular values of the outcome variable occurring; in this case the probability that multiple, single or no tuberculous possums will be found at a site, given the nature of the habitat as represented through the values of the

predictor variables. We discuss both applications of the logistic regression model in this paper.

Slope of the land was negatively associated with both multiple and single TB-positive categories, indicating that TB possums were more likely to be caught on flatter land (Table 6, Table 7, and Table 8; Figure 2). A possible explanation for the association between multiple TB possums and flatter slopes found in this study may be that higher densities of possums occurred on flatter land because the habitat provided more favourable denning areas, and the higher density resulted in a higher contact rate amongst possums leading to more opportunities for the transmission of TB. Higher densities of possums have been found on flatter slopes in studies in two areas of New Zealand (Coleman et al., 1980; Clout and Gaze, 1984). At the Castlepoint study site the highest population density, measured by trap catch, was located in the northern part of the study site which was dominated by a very steep slope with a plateau area above the steep section (Pfeiffer, 1994). Clusters of dens used by TB possums were located on both very steep and flat slopes within this area. Both the high density of possums and the clustering of TB may have been associated with this part of the study area as it contained the habitat in which a high proportion of possum dens were located (Paterson, et al., 1995). In the 1975 Hohonu survey the majority of TB possums were caught in podocarp forest on flatter land. However, the infected possums that were found above the lowland podocarp forest were close to or on the crest of warm ridges (Cook, 1975). The results of these different studies indicate that the steepness of slopes may not be consistently associated with the location of TB possums and the location of higher possum densities may vary between different geographic locations. This may occur as a result of variability in environmental conditions that produce habitat that contains possum dens which favour TB transmission.

We considered possible biases in the data, that could have spuriously given rise to the significant effect of slope in our study. At some sites the traps had either been laid at the bottom of a range of hills or at the top of a steep gully, where the land extending away from the traps was flat for up to 100 metres then became steep. At these locations the habitat at the trap site may have been different to that in which the possums regularly denned. However, at a larger number of sites the slope of the terrain at the trap site continued for a long distance into the area beyond the study site, thus we could have more confidence that the habitat at the trap site represented that in which the possums denned. We considered the possibility that TB possums were more likely to be trapped on flatter areas because they were too weak to reach

their original denning area, which is where they may have been infected. Paterson (1993) reported this phenomenon in the final stage of the disease. However, Jackson et al. (1995a) found that there was no difference in live weight and body condition for possums with TB lesions and those without, until the final stages of the disease. Given that the data for this study was sourced from cross-sectional surveys that had been conducted over many different seasons of the year and over many different years, possums were likely to have been caught at many different stages of the disease. It is thus unlikely that sufficient TB possums were in the final stages of disease to bias the results in this way.

A comparison of the final models for multiple and single TB-positive sites (Table 6, Table 7 and Table 8) shows that den factors were important in differentiating multiple TB-positive from negative sites but not in differentiating single TB-positive from negative sites. This supports the hypothesis generated from previous studies that possum TB clusters are associated with possum denning areas (Pfeiffer, 1994; Jackson, 1995). Furthermore, if we consider the classification of multiple TB-positive sites used in this study to represent persistent TB clusters and single TB-positive sites to represent sporadic TB clusters as proposed by McKenzie and Morris (1995), the results of this study suggest that factors influencing the location of persistent clusters are more likely to be associated with the quality of dens than those influencing the location of sporadic clusters. We acknowledge that it is difficult to accurately classify sites as persistent or sporadic TB clusters purely by the number of TB possums caught during a cross-sectional survey. However, given the relatively low success in catching TB possums in surveys due to clustering of these possums, single and multiple were reasonable proxies for persistent clusters and sporadic clusters, although this relationship has not been tested.

In the model that included the abundance of dens of each quality, multiple TB-positive sites were positively associated with the abundance of quality 1 dens. The quality 1 dens recorded in this study were located above ground in the crowns of tree ferns or in large trees which either contained epiphytes or had large branches with flat areas on which possums could comfortably sit. In the introduction we made the observation that areas in which a higher proportion of possums denned in trees above ground level may have a lower risk of supporting possum TB clusters, as the higher proportion of den sites above ground did not provide an environment conducive to the transmission of TB. The positive association between abundance of quality 1 dens and multiple TB-positive sites found in our study does not appear to be consistent with this hypothesis. A possible explanation is that the dens

recorded in this study were not located by radio tracking possums. All potential den sites were recorded regardless of whether they were used by possums or not, thus a number of the dens that were recorded may never have been used by possums. Quality 1 dens in this study may have indirectly represented other habitat factors associated with the quality of dens, with a higher abundance of quality 1 dens possibly indicating sites with a higher abundance of favourable possum dens of other quality scores.

In the model that represented den factors by the presence/absence of multiple dens of each quality, the presence of multiple quality 4 dens had a very strong positive association with multiple TB-positive sites (O.R. = 4) (Table 7). Quality 4 dens recorded in this study were most commonly located within the trunks or under the roots of large trees, or in logs or stumps that were often found at the same locations as large trees (Figure 6). One possible explanation for this association is that simultaneous den sharing may be more likely to occur in quality 4 dens, providing an extremely favourable opportunity for the transmission of TB due to the close contact of possums within an enclosed air space. In a study of the denning behaviour of female possums conducted near Dunedin, Cole (1993) found that simultaneous sharing occurred in 10% of the 40 core dens that she investigated, and that possums were more likely to share dens of higher quality. While a number of denning studies have found a small but significant amount of simultaneous sharing of dens by possums in the New Zealand environment (Green and Coleman, 1987; Fairweather, et al., 1987; Cole, 1993; Caley et al., 1998), den sharing is unlikely to be the only mechanism for transmission of TB amongst possums within the denning area. Den sharing was rarely observed at the Castlepoint study site (Pfeiffer, 1994; Paterson et al., 1995), yet TB has been maintained continuously at this site for over a decade (Corner, pers comm 1999). Furthermore, Caley et al. (1998) showed that den sharing is related to the density of the possum population. He found that the estimated probability of a denning event being shared reduced from 0.063 in an uncontrolled population with estimated density of 18 possums per hectare, to 0.012 when the population was reduced by 40%, and zero when reduced by 53%. Thus, den sharing may be less likely to explain the maintenance of TB infection in possum populations whose density has been greatly reduced by population control measures. While simultaneous den sharing has been found to be quite uncommon at the Castlepoint study site, in the study reported here areas where multiple TB possums had been trapped had a higher abundance of partially and fully enclosed dens than areas surrounding traps that caught only one or no tuberculous possums. This indicates that den quality is an important factor, though not exclusively by its influence

on den sharing.

Jackson (1995) and Pfeiffer (1994) hypothesized that social interaction around the denning area, either agonistic behaviour or grooming and mating, was another mechanism for transmission of TB. We propose a refinement of this hypothesis that there is a higher rate of agonistic encounters between possums during the process of selecting a den in areas where there is a high density of favourable possum dens. While this may sound counter-intuitive, the higher density of favourable dens may lead to a higher density of possums whose dens are located near to each other. Paterson (1995) found that the dens used by most possums were clustered in a small area at one end of the home range, where interactions with neighbours would be common. The rate of agonistic encounters may be even higher if there are a few very favourable dens within an area of high possum density, and where there is considerable competition between possums for these dens. While possums have not been found to defend an exclusive home range, a good den is a valuable resource and this appears to be an area of the home range that a possum is prepared to defend (Winter, 1976). Possums have been shown to use a number of different dens and many dens are used by a sequence of possums on different occasions (Brockie et al., 1989; Cowan, 1989; Cole, 1993; Pfeiffer, 1994; Paterson et al., 1995). Cole found that whilst possums in a forested area used up to 16 dens each over a 4 month period, they spent an average of 76% of their time in their three most popular den trees. Thus there may be more competition for certain dens, leading to more agonistic encounters in the denning area. Paterson (1993) and Winter (1976) both noted increased vocalisations whilst possums were returning to their dens before sunrise, which may indicate increased interaction whilst defending a den site.

Quality 2 dens were strongly negatively associated with multiple TB-positive sites when represented both as total abundance and as the presence or absence of multiple quality 2 dens (Table 6, Table 7 and Table 8). Figure 6 shows that a high proportion of quality 2 dens recorded in this study were located under logs, under piles of fallen branches and leaves, in recesses located in banks on steep slopes, or under ground cover such as long grass, fern, Carex species and bracken, with a lower proportion located in or under trees and stumps, or in hollow logs. Cole (1993) found that quality 2 and quality 4 dens were the most frequently used dens in a study that was conducted in an area of remnant forest in Otago, suggesting that possums find quality 2 and quality 4 dens more favourable than quality 1 or quality 3 dens. However, the environment in which quality 2 dens occur may significantly influence the degree to which possums find them favourable. For example, quality 2 dens occurring within

forest may be more likely to be located in association with large trees and/or the dead trunks or stumps of large trees or fallen logs, and possums may find quality 2 dens in these environments more favourable than in environments where dens of this quality are located under piles of fallen branches and leaf litter or under dense ground-level vegetation such as fern, grass or *Carex* species. As the dens recorded in this study were not located by radio tracking possums, all potential den sites were recorded regardless of whether they were used by possums or not. As a result, a number of dens that were recorded may never have been used by possums. The negative association of quality 2 dens found in this study may possibly be explained by the fact that at sites where there was a high abundance of quality 2 dens, these dens tended to be located under ground-level vegetation or piles of fallen branches and leaf litter which may not have been favourable possum denning areas. As a result there may have been a low density of possums denning in these areas, resulting in fewer opportunities for the transmission of TB between possums.

The two logistic regression models for "multiple TB-positive" sites produced in this study indicate that den characteristics are important in explaining the difference between multiple TB-positive and TB-negative sites. The two models are so close in R^2 (Table 6 and Table 7) that we cannot differentiate one as having significantly greater explanatory value than the other, and the two models can be considered to be interchangeable.

The mean top height of vegetation and percent of the ground covered with vascular vegetation were both positively associated with the odds of a single TB site, while percent coverage of tier 5 was negatively associated (Table 8). The positive association of ground-level vegetation found in this study reflects the fact that a high number of single TB sites included a single tree or a small group of trees growing on open pasture, whereas few negative sites included such cover. The negative association of the abundance of vegetation in tier 5, which included a height range of 0.3 to 2 metres, may relate to the same habitat factors as those associated with the negative relationship between quality 2 dens and multiple TB-positive sites; that is, TB possums were less likely to be caught at sites which had a high abundance of cover in the height range of 0.3 to 2 metres. Such locations may provide fewer favourable den sites for possums and may be more difficult for possums to move through because of the density of cover towards ground level, resulting in a lower density of possums compared with sites that have less cover at this height.

The logistic regression model may be used for predictive purposes to help target possum control efforts. A ROC curve was used to identify the cut-off probability for each model's

output that optimised the sensitivity and specificity with which sites were classified as multiple, single or negative. For the multiple TB model, a cut-off probability of 0.26 maximised the sensitivity (0.92) and the specificity (0.74), while a cut-off probability of 0.32 maximised the sensitivity (0.82) and the specificity (0.70) for the single TB model. The multiple TB-positive model has a high sensitivity of 0.92 which is associated with a lower specificity of 0.74. This lower specificity may be (in part at least) attributable to the fact that possum TB clusters do not expand to completely and permanently fill the habitat that has characteristics that favour cluster formation. This leaves large areas of potentially favourable habitat free of TB clusters. In addition, the temporal variation in the prevalence of TB results in some clusters not being detected by means of cross-sectional surveys. We do not have sufficient information to identify in which parts of favourable habitat the clusters will be located, hence we need to consider all such habitat as potentially containing a TB cluster. Some such locations will be "susceptible" and others "infected". A susceptible location is potentially just as dangerous as an infected one because if a TB-infected juvenile establishes a home range there, infection is likely to be spread and a persistent focus initiated.

The information resulting from this study could be used to derive maps describing the probability of the presence of tuberculous possum using a geographic information system (GIS). This would enable the possum TB risk of habitat patches to be assessed over large geographical areas, making this a useful planning tool that could facilitate the differential targeting of possum control resources based on TB risk. Digital habitat data with directly recorded detail of den quality and den density derived by remote sensing is currently unavailable, and is unlikely to be available in the future as it is difficult to determine this directly from remotely sensed data, and it is impractical to collect on-the-ground data over large areas. Thus den quality and density will need to be indirectly measured by other habitat factors such as plant species and vegetation structure (height and density) which act as proxies for den characteristics and possum density. This is likely to be less specific and possibly less sensitive in identifying areas with multiple high quality dens. However, it could be used as an indicator or screening test which can be supplemented by other information to more accurately identify the TB risk associated with particular habitat. Additional information could include cattle testing information, local farmer knowledge of possum TB hot spots, screening of habitat by farmers or control staff, and results of previous wildlife surveys.

More than 20 years ago, Cook (1975) made the statement that "even though a possum population may be reduced drastically by ..... well organised control operations, ...individuals may still be present to maintain a reservoir of infection menacing both cattle and immigrant possums. This inference could lead to a different phase of control that would start after routine operations have been completed." To date, the major focus of possum control efforts for TB control purposes has been on large-scale reduction of possum populations, and little emphasis has been put into the second phase, described by Cook, of identifying spatial reservoirs of infection. As the populations of possums in VRAs are reduced, control needs to focus more intently on identifying areas where tuberculous possums are most likely to remain and from which the infection could rebuild in an area. This study has provided insights into the epidemiology of TB in possums and information on habitat factors that could facilitate the implementation of a second phase of possum control that is targeted at identifying remaining reservoirs of tuberculous possums. Such an approach is likely to increase the effectiveness of possum control programmes in eradicating TB from possum populations.

Bibliography

Allen, R.B. (1992). RECCE. An inventory method for describing New Zealand vegetation. FRI Bulletin No. 181. Ministry of Forestry, New Zealand.

Anon. (1986). Epidemiological studies of tuberculosis in cattle and possums. In: S. Hennessy (Ed), Surveillance 13 (3): 20-24.

Barlow, N.D. and Kean, J.M. (1995). Spatial models for tuberculosis in possums. In: F. Griffin, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. Proceedings of the Second International Conference on *Mycobacterium bovis*. Otago Conference Series No. 3, University of Otago Press, Dunedin, New Zealand. 214-18.

Bian, L. and West, E. (1997). GIS modelling of elk calving habitat in a prairie environment with statistics. Photogrammetric Engineering and Remote Sensing 63 (2): 161-67.

Brockie, R.E., Hearfield, M.E., White, A.J., Waddington, D.C., Hay, J.R. (1987). Bovine tuberculosis in a possum from the Orongorongo Valley, Wellington. New Zealand Veterinary Journal 35: 201-203.

Brockie, R.E., Herritty, P., Ward, G. W., and Fairweather, A. A. C. (1989). Population study on Hawkes Bay farmland possums. DSIR Ecology Division Report.

Brockie, R.E. (1991). Ecology of an uninfected farm possum population. In: R. Jackson (Convenor), Symposium on Tuberculosis. Foundation for Continuing Education of the N.Z.V.A., No.132, Palmerston North, New Zealand. 53-66.

Broschart, M.R., Johnston, C.A., and Naiman, R.J. (1989). Predicting beaver colony density in boreal landscapes. Journal of Wildlife Management 53 (4): 929-34.

Buckland, S.T. and Elston, D.A. (1993). Empirical models for the spatial distribution of wildlife. Journal of Applied Ecology 30: 478-95.

Caley, P. (1995a). Effects of possum maintenance control and ecological factors on levels of bovine tuberculosis in cattle and possums. Lancare Research Contract Report LC9495/120, Landcare Research, Palmerston North.

Caley, P. (1996). Is the spatial distribution of tuberculous possums influenced by den "quality"? New Zealand Veterinary Journal 44 (5): 175-78.

Caley, P., Spencer, N.J., Cole, R.A., and Efford, M.G. (1998). The effect of manipulating population density on the probability of den-sharing among common brushtail possums, and the implications for transmission of bovine tuberculosis. New Zealand Veterinary Journal 25: 383-92.

Clout, M.N. and Gaze, P.D. (1984). Brushtail possums (Trichosurus vulpecula kerr) in a New Zealand beech (Nothofagus) forest. New Zealand Journal of Ecology 7: 147-55.

Cole, R. (1993). Denning behaviour of female brushtail possums (Trichosurus vulpecula) at Pigeon Flat, Otago. University of Otago Wildlife Management Report Number 41, Dunedin, New Zealand.

Coleman, J.D. (1988). Distribution, prevalence, and epidemiology of bovine tuberculosis in brushtail possums, Trichosurus vulpecula, in the Hohonu Range, New Zealand. Australian Wildlife Research 15: 651-63.

Coleman, J.D., Gillman, A., and Green, W.Q. (1980). Forest patterns and possum densities within podocarp/mixed hardwood forests on Mt Bryan O'Lynn, Westland. New Zealand Journal of Ecology 3: 69-84.

Coleman, J.D., Drew, K. W., and Coleman, M. C. (1992). Prevalence and spatial distribution of bovine tuberculosis in a possum population, Ahaura Valley, Westland. Within-forest patterns: December 1992. Landcare Research Contract Report: LC9293/90. Manaaki-Whenua - Landcare Research, Christchurch, New Zealand.

Coleman, J.D., Jackson, R., Cooke, M.M., and Grueber, L. (1994). Prevalence and spatial distribution of bovine tuberculosis in brushtail possums on a forest-scrub margin. New Zealand Veterinary Journal 42 (4): 128-32.

Coleman, J.D., Coleman, M. C., and Cooke, M. M. (1996). Prevalence and spatial distribution of bovine tuberculosis in a possum population Ahaura Valley, Westland. Forest-edge patterns: Year 4 - August 1995. Landcare Research Contract Report: LC9596/66. Landcare Research, Christchurch, New Zealand.

Coleman, J., Fraser, W., and Nugent, G. (1998). Optimal buffer widths for the control of TB in wildlife on forest/pasture margins. In: V. Anderson, and P. Livingstone (Eds), Workshop to discuss the possibility of eradicating TB from the northern central North Island by 2010. Animal Health Board, Wellington, New Zealand.

Cook, B.R. (1975). Tuberculosis in possums. Hohonu Mountain MAF/NZFS Project117. Animal Health Division, Ministry of Agriculture and Fisheries, Wellington, New Zealand.

Cowan, P.E. (1989). Denning habits of common brushtail possums, *Trichosurus vulpecula*, in New Zealand lowland forest. Australian Wildllife Research 16:63-78.

Davidson, R.M. (1976). The role of the opossum in spreading tuberculosis. New Zealand Journal of Agriculture 133: 21-25.

Fairweather, A.A.C., Brockie, R.E., Ward, G.D. (1987). Possums (*Trichosurus vulpecula*) sharing dens: a potential infection route for bovine tuberculosis. New Zealand Veterinary Journal 35: 15-16.

Green, W.Q. and Coleman, J.D. (1987). Den sites of possums, *Trichosurus vulpecula*, and frequency of use in mixed hardwood forest in Westland, New Zealand. Australian Journal of Wildlife Research 14: 285-92.

Hall, G.M.J. (1992) PC-RECCE. Vegetation inventory data analysis. FRI Bulletin No. 182. Ministry of Forestry, New Zealand.

Hickling, G.J. (1989). Assessment of possum control in an area of endemic bovine tuberculosis. Forest Research Institute, Christchurch, New Zealand.

Hickling, G.J. (1991). Ecological aspects of endemic bovine tuberculosis infection in brushtail possum populations in New Zealand. In: Proceedings of the 9th Australian Vertebrate Pest Control Conference. 332-36.

Hickling, G. (1995). Clustering of tuberculosis infection in brushtail possum populations: implications for epidemiological simulation models. In: F. Griffin, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. University of Otago Press, Dunedin, New Zealand. 174-77.

Hickling, G.J. and Efford, M.G. (1996). Assessing the risk of bovine tuberculosis infection for New Zealand cattle herds in TB vector risk areas (VRAs). Statistics in Ecology and Environmental Monitoring 2: 79-90.

Jackson, R. (1995). Transmission of tuberculosis (*Mycobacterium bovis*) by possums. Unpublished PhD thesis, Massey University, New Zealand.

Jackson, R., Cooke, M.M., Coleman, J.D., and Morris, R.S. (1995a). Naturally occurring tuberculosis caused by *Mycobacterium bovis* in brushtail possums (*Trichosurus vulpecula*): I. An epidemiological analysis of lesion distribution. New Zealand Veterinary Journal 43 (7): 306-14.

Jackson, R., Cooke, M.M., Coleman, J.D., Morris, R.S., deLisle, G.W., and Yates, G.F. (1995b). Naturally occurring tuberculosis caused by *Mycobacterium bovis* in brushtail possums (*Trichosurus vulpecula*): III. Routes of infection and excretion. New Zealand Veterinary Journal 43 (7): 322-27.

Kent M., and Coker, P. (1992). Vegetation description and analysis. A practical approach. Belhaven Press, London.

Langmuir, A.D. (1961). Epidemiology of airborne infection. Bact. Rev. 25: 173-81.

Lindenmayer, D.B., Cunningham, R.B., Tanton, M.T., Smith, A.P., and Nix, H.A. (1990). Habitat requirements of the mountain brushtail possum and the greater glider in the montane ash-type eucalypt forests of the central highlands of Victoria. Australian Journal of Wildlife Research 17: 467-78.

Livingstone, P.G. (1997). Update on the New Zealand TB situation. In: Proceedings of a Seminar on Possum and Mustelid Control Research. National Possum Control Agencies, Wellington, New Zealand. 17-30.

Lugton, I. W. (1997). The Contribution of Wild Animals to the Epidemiology of Tuberculosis in New Zealand. Unpublished PhD thesis, Massey University, Palmerston North, New Zealand.

McKenzie, J.S.and Morris, R.S. (1995). Development of spatial predictors of tuberculosis infection risk in wildlife populations and their application within a decision support framework. In: F. Griffen, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. Proceedings of the Second International Conference on Mycobacterium bovis. University of Otago Press, Dunedin, New Zealand. 168-70.

Morris, R.S., Pfeiffer, D.U., and Jackson, R. (1994). The epidemiology *of Mycobacterium bovis* infections. Veterinary Microbiology 40 (1): 153-77.

O'Neil, B.D. and Pharo, H.J. (1995). The control of bovine tuberculosis in New Zealand. New Zealand Veterinary Journal 43 (7): 249-55.

Pannett, G. (1995). Possum control in the Wellington region: how successful has it been? In: F. Griffen, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. Proceedings of the Second International Conference on Mycobacterium bovis. Otago Conference Series 3, University of Otago Press, Dunedin, New Zealand. 294-96.

Paterson, B. M. (1993). Behavioural patterns of possum and cattle which may facilitate the transmission of tuberculosis. Masters thesis, Massey University, Palmerston North.

Paterson, B.M., Morris, R.S., Weston, J., and Cowan, P. (1995). Foraging and denning pattern of brushtail possums and their possible relationships to contact with cattle and tuberculosis transmission. New Zealand Veterinary Journal 43 (7): 281-88.

Pereira, J.M.C. and Itami, R.M. (1991). GIS-based habitat modelling using logistic multiple regression: a study of the Mt. Graham red squirrel. Photogrammetric Engineering and Remote Sensing 57 (11): 1475-86.

Pfeiffer, D. U. (1994). The role of a wildlife reservoir in the epidemiology of bovine tuberculosis. Unpublished PhD thesis, Massey University, New Zealand.

Pfeiffer, D.U., Hickling, G.J., Morris, R.S., Patterson, K.P., Ryan, T.J., and Crews K.B. (1995). The epidemiology of *Mycobacterium bovis* infection in brushtail possums

(*Trichosurus vulpecula* Kerr) in the Hauhungaroa Ranges, New Zealand. New Zealand Veterinary Journal 43 (7): 271-80.

Ward, G.D. (1978). Habitat use and home range of radio-tagged opossums *Trichosurus vulpecula* (Kerr) in New Zealand lowland forest. In G.G. Montgomery (Ed), The ecology of arboreal folivores. Proceedings of the national zoological park symposium. No. 1. Smithsonian Institution Press, Washington D.C. 267-287

Winter, J. (1976). Behaviour and Social Organisation of the brush-tail possum. Unpublished PhD thesis, University of Queensland, Australia.

CHAPTER 3

Spatial and temporal patterns of vector-borne tuberculosis infection in beef breeding cattle[2]

Abstract

We used the spatial and space-time scan statistics to identify clustering of three measures of tuberculosis infection (TB) on beef breeding farms, in an area of New Zealand where TB is endemic in the wild possum population. We found significant spatial clusters for the disease frequency measures "five-year cumulative incidence" (1986-90) and "percent TB-positive years", and significant space-time clusters for "annual cumulative incidence". An extremely useful feature of these scan statistics is their geographic output, which enables the user to map the locations of significant clusters. This enabled us to develop greater insights into factors associated with the spatial and temporal distribution of TB in the study area. We conclude with a discussion of the application of these methods as a surveillance tool in the TB management programme, to detect both high-rate and low-rate TB incidence clusters at the farm level.

Introduction

Tuberculosis (TB) due to infection with *Mycobacterium bovis* is endemic in brushtail possums (*Trichosurus vulpecula*) in areas, known as vector risk areas (VRAs), which cover an estimated 24% of New Zealand. Possums are believed to be the source of TB infection for the majority of TB cases in cattle and deer in these areas (Livingstone, 1997). Patterns of the distribution of TB within infected possum populations have been explored in some detail, with evidence of clustering of TB around possum denning areas, covering a cross-sectional area of the order of 40 meters (Pfeiffer, 1994; Hickling, 1995). There is evidence that some possum TB clusters, commonly referred to as TB hot spots, persist at the same location for many years while others are more sporadic and persist for short periods of time only (McKenzie et al., 1997). The location of both persistent and sporadic clusters has been shown to be associated with specific habitat factors that influence the possum denning environment (see Chapter 2). Detailed investigations of spatial patterns of TB at a larger scale within VRAs have not been conducted. Veterinarians responsible for implementing the TB control programme in New Zealand have observed clustering of TB at the farm level within VRAs, with farms in some areas having a higher incidence than others. Quantitative analysis of the spatial patterns of TB at the farm level would enable us to identify if the observed clustering is likely to occur due to chance alone, and would help us develop hypotheses about risk factors that are associated with vector-borne TB in cattle at the farm level.
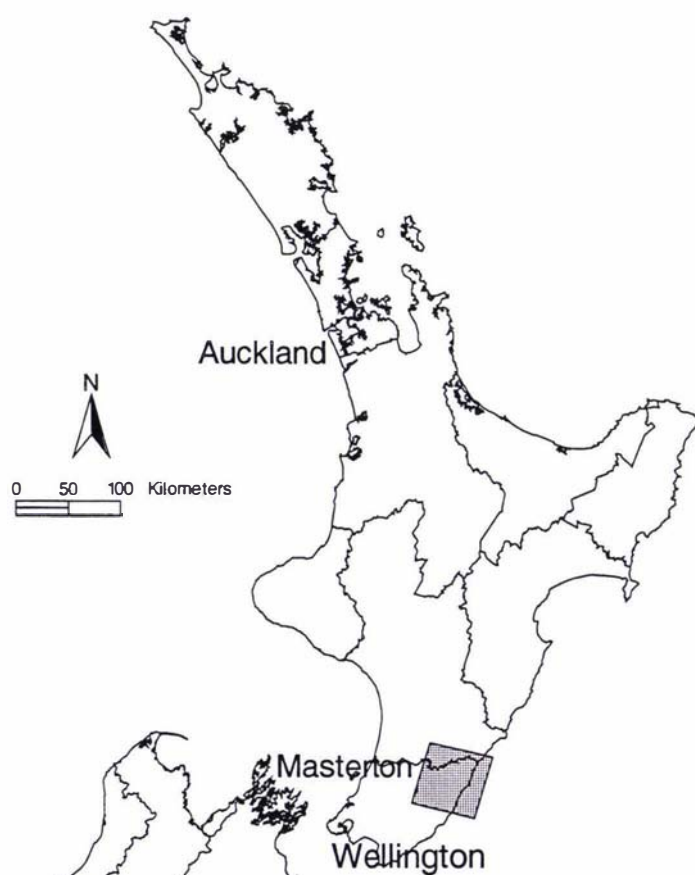
Figure 8. Map of the North Island of New Zealand showing the location of the study area.

The recent development of a digital geographically referenced national farm database for New Zealand, known as Agribase (Sanson and Pearson, 1997), has made it possible to map farm boundaries within a geographic information system (GIS), and to conduct spatial analyses using this data. The development of this database has coincided with the recent development of software (SaTScan) that can test for the significance of clustering within space and/or time within a geographic data set, and also identify the location of the clusters (Kulldorff and Nagarwalla, 1995; Hjalmars et al., 1996; Kulldorff et al., 1997). Most tests for spatial clustering of health events, test for global clustering throughout the study region without the ability to pinpoint the location of specific clusters (Kulldorff et al., 1997a). Likewise available space-time interaction tests evaluate global space-time interaction throughout the study region and time period without the ability to pinpoint the location of specific clusters. The spatial scan statistic and the space-time scan statistic used in SaTScan have the advantage of being able to detect the location and size of clusters and to test the significance of these clusters.

The driving force for the development of this methodology has been to develop a tool to help health officials evaluate local disease cluster alarms, where members of the public are concerned about a perceived increase in the incidence of disease in a localised area (Hjalmars et al., 1996; Kulldorff et al., 1998). This software helps determine whether the cluster has occurred by chance or whether the excess is so great that it is probably due to some common elevated risk factor acting at this location. The advantage of this method is that it tests clusters of all possible spatial and space-time sizes without any pre-selection bias as to the area within which the perceived cluster is located. We were interested in applying this methodology to vector-borne TB in cattle in an area in New Zealand where possums are known to be infected with TB, particularly because of the scan statistic's ability to identify the location of individual clusters.

In this paper we describe the application of the spatial scan statistic and the space-time scan statistic to TB incidence data in beef breeding farms in a VRA. We compare spatial and spatio-temporal patterns of three different TB measures at the farm level, and discuss hypotheses regarding the association of these patterns with specific risk factors.

Materials and Methods

Study area

The study farms were located in a 60 kilometer square area in the north-eastern part of the Wellington region, known as the Wairarapa (Figure 8). The study area corresponded with the coverage of a SPOT3 multispectral satellite image that was used to generate a vegetation map for a related study (McKenzie and Dymond, in preparation). This is a pastoral farming area dominated by sheep and beef production where TB has been endemic in cattle since the 1950s (Shortridge, 1981). The traditional test and slaughter approach was successful in reducing the incidence of TB in cattle in this area until the late 1960s, when epidemiological evidence in this area and other parts of the country suggested that possums were acting as a reservoir for TB. TB was first found in possums in the Wairarapa in 1968 (Anon, 1993). Since that time tuberculous possums have been caught at over 140 different locations in the Wairarapa. Other feral and wild animals infected with TB have also been caught at several locations throughout the region since the early 1970s.

Large scale possum population control programmes were implemented in the eastern and southern zones of the Wairarapa between 1976 and 1980 (Anon, 1993). TB cattle reactor

numbers fell by 80% within two years of the initial control efforts. However, after 5 years they began to increase and were at pre-control levels 7-8 years after the original operation. During the period 1980-1988 there were only small areas under possum control and this period was associated with a steady increase in the number of TB reactors throughout the area. Since 1989 the area of land under possum control has increased steadily with a total of 100,000 hectares being treated in 1996/97. The increased control efforts have been associated with a steady reduction in the annual incidence of TB reactors (Livingstone, 1997).

Cattle TB data

We obtained cattle and deer TB testing data from the National Livestock Database (NLDB) which is administered by AgriQuality New Zealand (the state veterinary service) (Ryan, 1997). This database is designed to manage and record TB testing on farms as required by the national TB control programme. TB data are recorded in the NLDB on a herd basis and records for the Wairarapa date back to 1979. In 1996 the NLDB was amended to enable herds to be linked to the farm on which they were managed by the addition of a unique farm identification number. This was made possible by the development of a national geographically referenced farm database, known as Agribase, in which each farm is assigned a unique farm identification number (see below for details). This database has provided a vital link between TB history data and the geographic units from which the data arises, enabling mapping and spatial analysis of TB data to be conducted at the farm level.

The year in which a possum control operation was implemented by the Regional Council is recorded for each farm in the NLDB. The date recorded is September 1st for the test year during which the operation was conducted. Most operations were conducted between January and June of the calendar year following that recorded in the database. We thus assumed that a possum control operation would begin to have an influence on the cattle TB testing results in the financial year following that during which the vector control programme was recorded as taking place.

Farm boundary map

A map of farm boundaries for the study area was obtained from Agribase, a national geographically referenced database of farms in New Zealand that has been developed by AgriQuality New Zealand (Sanson and Pearson, 1997). Farm boundaries have been built up from the digital cadastral database (DCDB). Each land parcel in the DCDB that has been matched to a farm has been allocated the appropriate farm identification number. The

adjoining boundaries of adjacent land parcels with the same identification number have been dissolved. However, most farms still comprise multiple polygons as the component land parcels are not immediately adjacent to each other. The database for this area is gradually being updated by AgriQuality NZ staff in the course of their routine activities. However, it was not complete at the time of conducting this study, which limited the number of farms that could be included in the analysis.

Study population

The study population comprised 129 beef breeding farms which each had an area greater than 10 hectares, and for which we had accurate farm boundary information. Only farms on which the cattle herd had been tested for a minimum of ten years up to and including the first year of possum control were included in the study, to maximise the chance that the full range of temporal variation in TB incidence was included for each study farm. Fourteen farms in the study group had a deer herd as well as a cattle herd on the property. Only cattle data were considered on these farms, as the deer were generally confined to a small specially fenced area of the farm and we made the assumption that their exclusion did not confound the cattle data.

Data availability and quality

Our data set of 129 farms represented a subset of farms in the study area, as this study of spatial patterns was part of a larger study to identify possum habitat and other geographic factors associated with TB incidence in cattle at the farm level. For this study, we required accurate farm boundary information to ensure accurate measurement of the geographic variables on farms, such as the area of different vegetation classes, length of rivers and distance from forests (see Chapter 5). Matching of all land parcels to farms had not been completed at the time of this study. As a result, we had to exclude a number of farms with inaccurate boundary information from the study. Other beef breeding farms were excluded as they did not have ten years of TB test data. Other farms or areas were excluded as they had no cattle, such as forests, or because they were not beef breeding management systems; for example, dairy or dry beef farms. This latter group of farms was excluded as the incidence of TB was more likely to be confounded by the movement of cattle between different farms, and was thus less likely to reflect possum TB conditions on the home farm.

Very high quality data is desirable for a spatial analysis as areas with missing data or data associated with inaccurate locations can generate effects of the same order of magnitude as
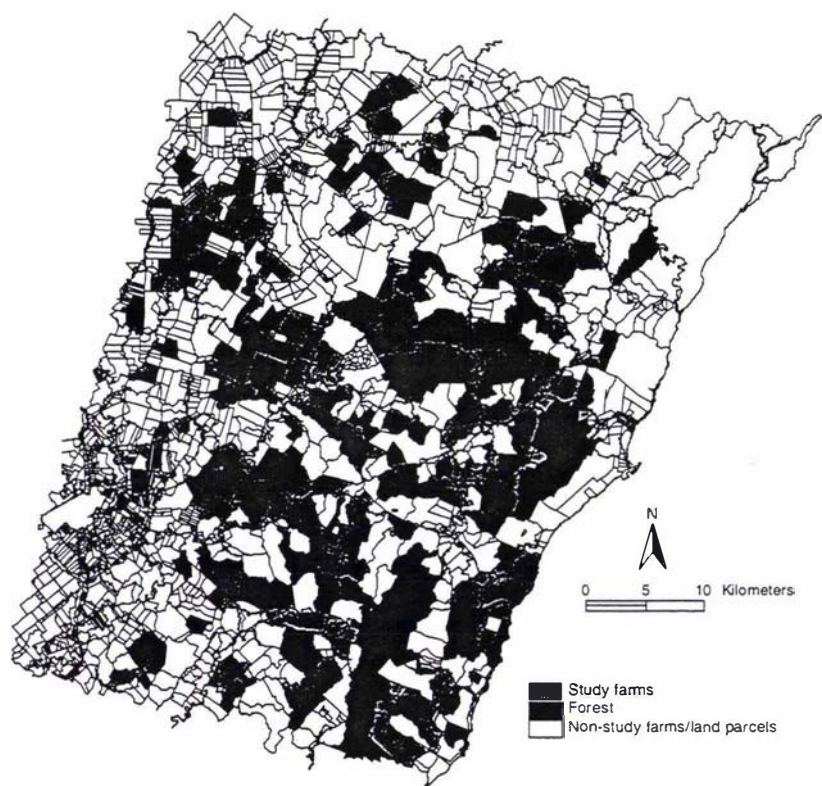
Figure 9. Distribution of the 129 beef breeding study farms.
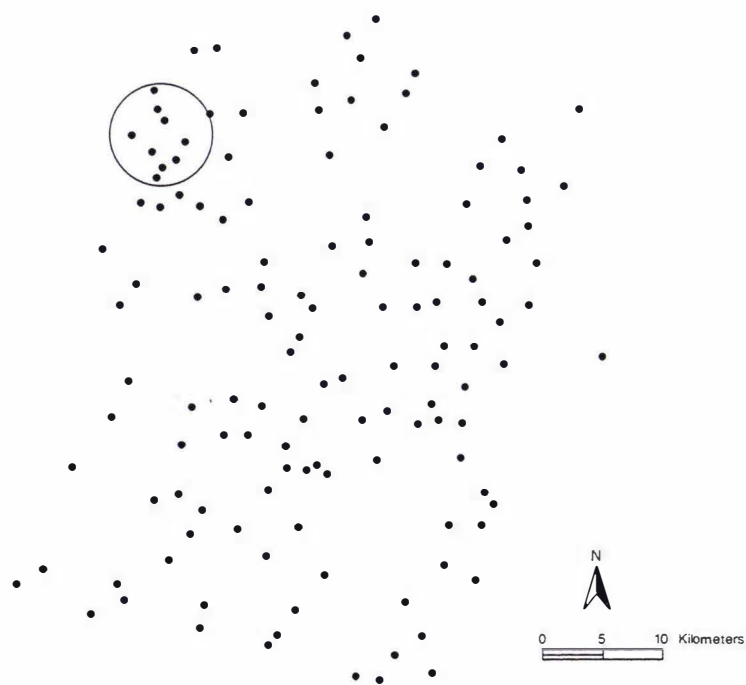
Figure 10. Distribution of point data used to represent the location of the study farms. Circle indicates an area where smaller farm size leads to a higher density of points.

might be expected from environmental influences (Walter, 1993). The distribution of farms used in this study is patchy (see Figure 9), and is not of a high quality for spatial analysis. However, we believe that the density of study farms is sufficiently high in the south-east through the central to the north-west part of the study area to demonstrate the application of the spatial analytical techniques described in this paper, and to draw some preliminary conclusions on factors associated with spatial patterns in these areas.

Measures of frequency of occurrence of cattle TB infection

For the purposes of this study, TB-positive cattle included all cattle that were recorded in the NLDB as being either caudal fold tuberculin test-positive or comparative cervical tuberculin test-positive, regardless of whether they had post mortem lesions consistent with TB. The caudal fold tuberculin test is the routine test used for detection of TB in cattle in New Zealand. The comparative cervical tuberculin test is used in some herds as a supplementary test if there is doubt about the specificity of the caudal fold test. Cattle testing positive were included as TB-positives regardless of the reason for the testing, which could have been annual herd tests, sale tests, miscellaneous tests, and tests associated with the movement of animals between farms. TB-positive cattle also included cattle that were slaughtered for market purposes and were recorded as having had TB lesions detected during routine inspection at the slaughter plant.

We used the New Zealand government's financial year (1 July – 30 June) as the annual time period for calculating TB frequency measures as this was consistent with the time period currently used by AgriQuality New Zealand for reporting TB data. Herds whose annual herd test was conducted as two or more partial herd tests in different financial years, had the results from that test allocated to the financial year in which the final part of the test was conducted.

Three different frequency measures of TB occurrence in cattle were calculated.

i) Annual cumulative incidence
Cumulative incidence (CI) was calculated for each herd for each financial year for which TB surveillance data was available in the NLDB within the period 1979-96. The numerator was the sum of all TB-positive cattle recorded during each financial year. The denominator was the average number of cattle tested in whole herd tests during the financial year. Test-eligible animals included all cattle on the farm over the age of 12 weeks. If a whole herd test was conducted as a series of part herd tests, then the number tested at each part herd test was summed to give a total for that whole herd test.

Cumulative incidence was chosen over incidence density or corrected annual incidence (corrected to 365 days) as all herds in the study area were under annual testing thus the length of time between whole herd tests did not vary significantly between herds. Zewdie (1997) showed that there was a high degree of correlation between TB cumulative incidence and incidence density in data for herds that were on an annual testing programme.

ii) Five-year cumulative incidence (1986-90)

A five-year CI was calculated for the period 1986-90. We chose this time period to eliminate the confounding effect of possum control programmes on TB incidence from 1990 onwards. The numerator of the five-year CI was the sum of all TB-positive cattle recorded during 1986-90, and the denominator was the average number of cattle tested in the annual herd test during this period.

iii) Percent years with TB-positive cattle

The numerator for this measure was the number of years when the herd had at least one TB-positive cattle beast. The denominator was the number of financial years that the herd had been included in AgriQuality New Zealand's TB surveillance programme, up to and including the first year during which a possum control operation was conducted on a farm.

Statistical analysis

Spatial clustering was evaluated for the five-year CI (1986-90) and percent TB-positive years using a spatial scan statistic (Kulldorff and Nagarwalla, 1995; Kulldorff et al., 1997). The method uses point data, which in this study was the centroid of the largest polygon of each study farm (see Figure 10). The spatial scan statistic applies a moving circular window that centres on each point in turn. At each position, the radius of the window varies so that the window includes a minimum of zero neighbouring points up to a maximum number set in the analysis. We set the maximum size of a window such that it included up to 50% of the study population. This avoids the situation whereby the spatial window becomes so large that it is inappropriate to refer to a cluster in that zone as representing any mechanisms of epidemiological relevance. During the analysis the algorithm generates a very large number of distinct circular windows, each with a different set of neighbouring farms in it, and each a possible candidate for containing a cluster of TB cattle cases or of TB-positive years. The scan statistic assumes that cases are Poisson distributed with uniform risk over space under the null hypothesis of absence of spatial clustering. For each window the method tests the null hypothesis against the alternative hypothesis that there is an elevated risk of TB in cattle or

TB-positive years within compared with outside the window. For each window the numbers of disease cases inside and outside the window are noted, together with the expected number of cases reflecting the population at risk. On the basis of these numbers, the likelihood function is calculated for each circular window. The window with the maximum likelihood is denoted the 'most likely cluster'. The expected distribution of the likelihood ratio test statistic under the null hypothesis is obtained by repeating the same analytical exercise for a large number (we used 999) of random replications of the data set through a Monte Carlo sampling process. It is then used to estimate a p-value expressing the probability of obtaining the observed cluster under the null hypothesis (Kulldorff and Nagarwalla, 1995).

In addition to the most likely cluster, the spatial scan statistic identifies secondary clusters in the data set and can order them according to their likelihood ratio. Secondary clusters are reported if their likelihood ratio is larger than the likelihood ratio for the most likely cluster for at least one data set simulated under the null hypothesis. The software does not report any sets of farms that partly overlap the most likely cluster and that have a likelihood almost as high, since most of them provide little additional information. As a result, we need to consider the results of the analysis as representing the approximate location of a cluster whose exact boundaries are uncertain.

Space-time clustering of annual TB incidence on farms was evaluated using the space-time scan statistic (Kulldorff et al., 1998). The space-time scan statistic is similar to the spatial scan statistic except that it is defined by a cylindrical window with a circular geographic base and with height corresponding to time. The height reflects any possible time interval up to a maximum set for the analysis. The window is then moved in space and time so that for each possible geographic location and size it also considers each possible time interval. Cases were assumed to be Poisson distributed with uniform risk over space and time under the null hypothesis, and with different risks inside and outside at least one of the cylinders under the alternative hypothesis. Space-time analyses were conducted at two different space-time scales. In the first analysis the maximum settings recommended for both space and time windows were used (Kulldorff and Nagarwalla, 1995). These were 50% of the study farms (65 farms) and 90% of the study years (16 years). In the second analysis the maximum extent of the spatial window was limited to 6 farms (5% of the total number of study farms), and the maximum time window to 2 years (12% of the available time). These limits were chosen as we were interested in identifying if there was evidence of space-time clustering on immediately neighbouring farms over consecutive years.
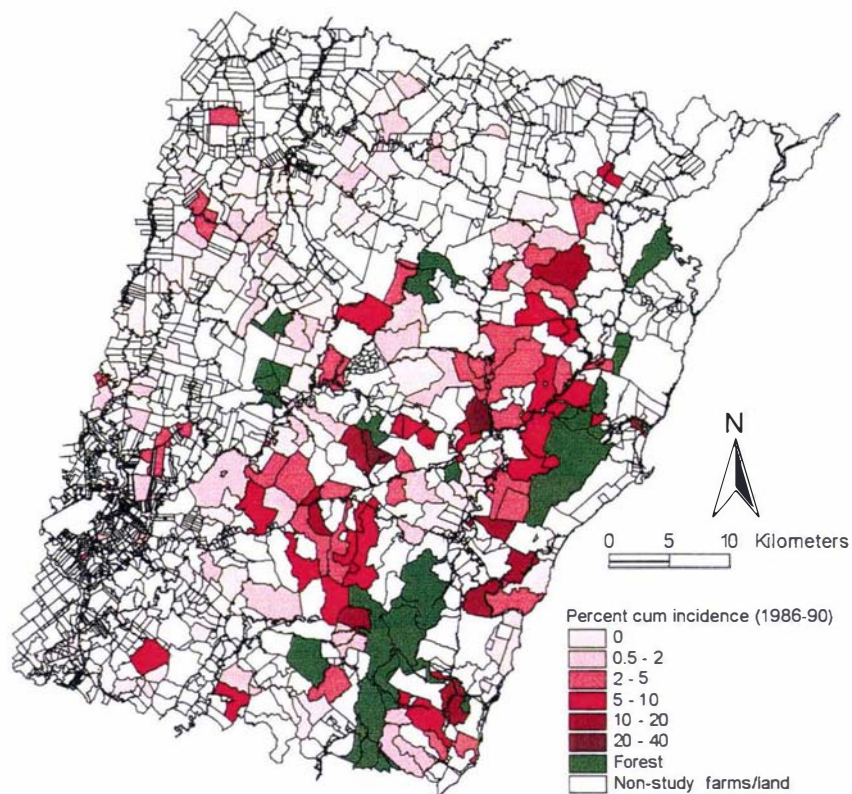
Figure 11. Distribution of the five-year cumulative incidence of TB-positive cattle (1986-90).
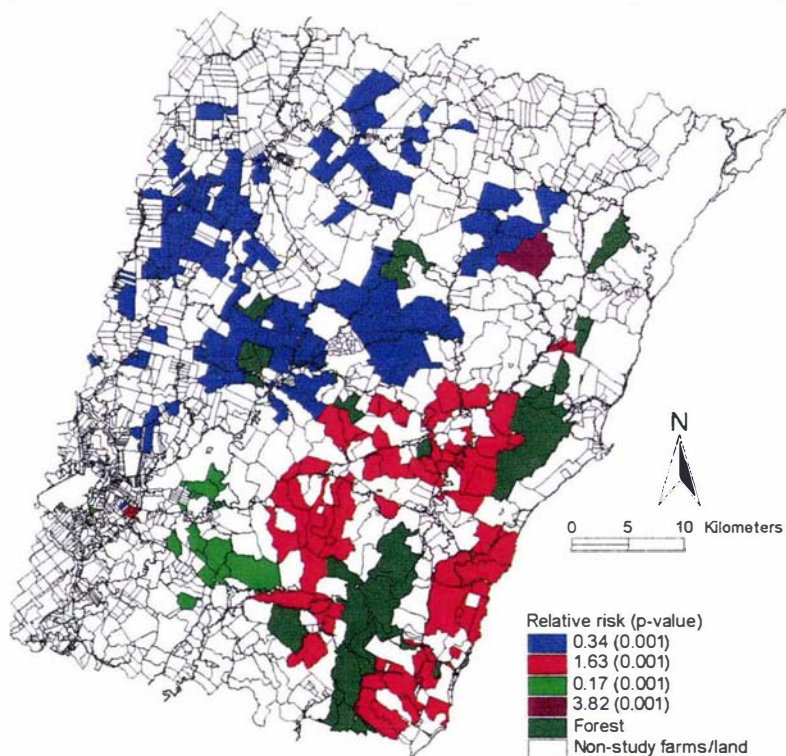
Figure 12. Location of the most likely high and low-rate clusters for five-year CI (1986-90), together with the relative risk of TB-positive cattle on farms within versus without each cluster.

All spatial, space-time and temporal clustering calculations were performed using SaTScan, which has been designed specifically to implement the spatial scan and the space-time scan statistics (Kulldorff et al., 1997b). SatScan produces a text file that lists the geographic coordinates and the identification of each farm included in each of the significant clusters. This data can be used in a GIS to map the locations of the clusters. The centroid of the largest polygon was used to represent a single point location of each farm in the analyses. However, the resulting clusters were mapped using the entire boundary of each farm whose centroid was within the cluster, as this enabled us to visualise the full extent of clustered farms. As some farms had blocks of land remote from the largest polygon that was used to obtain the centroid for the analysis, some polygons shaded as being associated with a cluster were located outside the main region of the cluster. Maps were produced using ArcView version 3.1 (ESRI, Redlands, California, USA).

Results

We show the distribution of the five-year CI (1986-90) for the 129 study farms in Figure 11, with median equal to 0.9% and a range of 0-39%. Results of the cluster analysis for the five-year CI using the spatial scan statistic showed that the most likely high-rate cluster included 39 farms located in the south-eastern part of the study area (Figure 12). The relative risk for TB-positive cattle occurring on farms within the cluster during 1986-90 compared with those outside the cluster was 1.63 (p=0.001). A secondary high-rate cluster included a single farm in the north-east that had a high relative risk of 3.83 (p=0.001) during the 1986-90 period. The most likely low-rate cluster for the five-year CI included 55 farms located in the north-western part of the study area (Figure 12). The relative risk for TB-positive cattle on farms within the cluster compared to those outside the cluster was 0.34 (p=0.001). A secondary low-rate cluster for CI included 8 farms located towards the south-western part of the study area where the relative risk of TB-positive cattle was 0.17 (p=0.001).

We show the distribution of the percent TB-positive years for the 129 study farms in Figure 13, with median equal to 37% and a range of 0-100%. The most likely high-rate cluster for percent TB-positive years included 66 farms located in the south-eastern part of the study area (Figure 14). The relative risk of a TB-positive year occurring on farms within the cluster compared to those outside the cluster was 1.47 (p=0.001). This was a large cluster that included just over half of the study farms. Two significant low-rate clusters were identified for percent TB-positive years. The most likely low-rate cluster included 29 farms in the north-
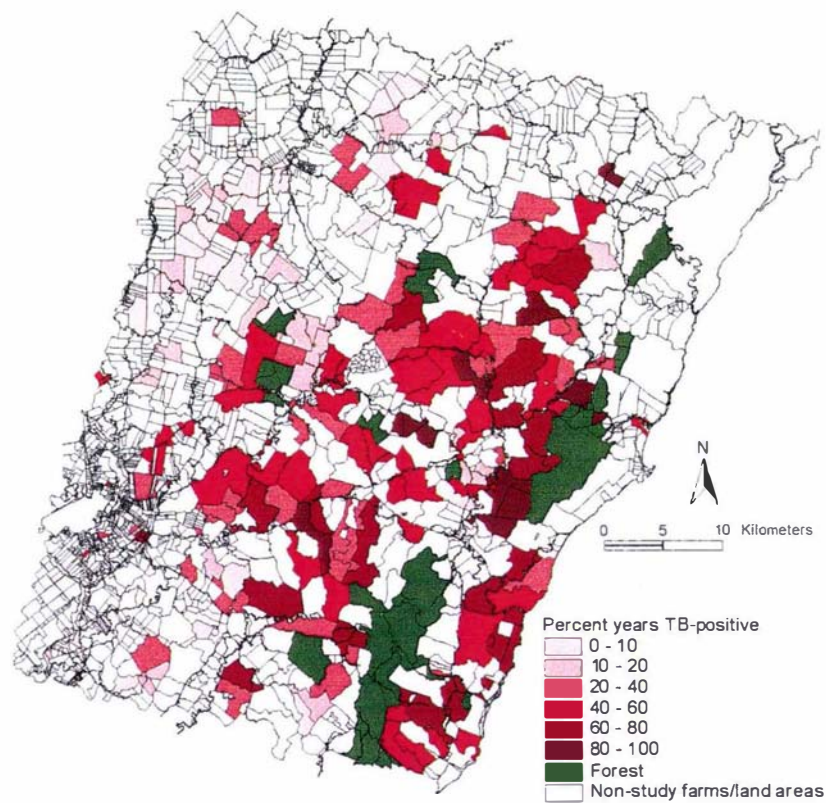
Figure 13. Distribution of percent TB-positive years.
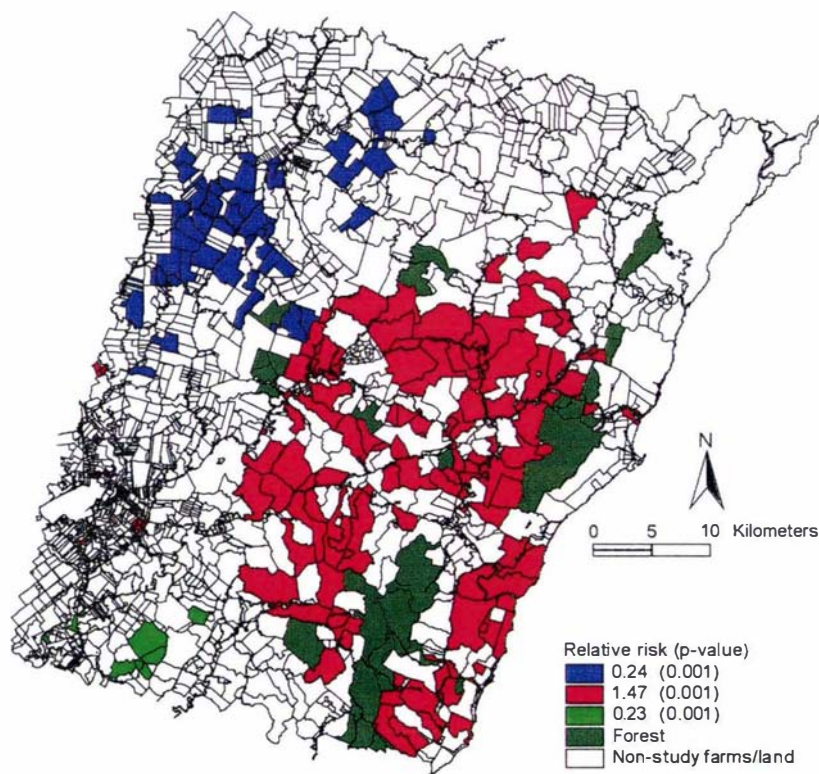
Figure 14. Location of the most likely high and low-rate clusters for percent TB-positive years.

western part of the study area with a relative risk of 0.24 (p=0.001) (Figure 14). The secondary low-rate cluster included 6 farms located in the south of the study area with a relative risk of 0.23 (p=0.001).

A comparison of the location and extent of the high-rate and low-rate spatial clusters for both the five-year CI and the percent TB-positive years in Figure 12 andFigure 14 shows that the high-rate clusters for both measures were located in the south-eastern part of the study area, and the low-rate clusters were both located in the north-western area. The high-rate cluster for percent TB-positive years included 66 farms and extended over a larger area than the high-rate cluster for the five-year CI, which included only 39 farms. A comparison of the 66 farms included in the high-rate cluster for percent TB-positive years with those in the five-year CI clusters, showed that 39 were included in the high-rate cluster for the five-year CI, 14 were included in the low-rate clusters for the five-year CI, and 13 were not included in any of the CI clusters. Conversely, the low-rate cluster for the five-year CI included 55 farms and extended over a larger area than that for percent TB-positive years which included only 29 farms.

We show the location of the seven significant space-time clusters from the analysis with a maximum time window of 90% and a maximum space window of 50% in Figure 15. Details of the years during which each cluster occurred together with its relative risk and p-value are shown in Table 9. The most likely space-time clustering occurred on a group of 27 farms located in the east of the study area during the period 1985-96. The relative risk of TB-positive cattle occurring on these farms during 1985-96 compared with any other space-time combination was 2.26 (p=0.001). A further two space-time clusters involved multiple farms; one included a group of 4 farms (cluster ID = 6) in the west in 1982 (relative risk = 4.77, p=0.003), and a second included 6 farms (cluster ID = 5) in the centre of the study area in 1979 (relative risk = 3.92, p=0.001). Four space-time clusters included a single farm in each, two of which had notably high relative risks for a one-year period. One farm (cluster ID = 2) had a relative risk of 22.2 (p=0.001) during 1995 and the other (cluster ID = 3) had a relative risk of 18.1 (p=0.001) during 1993 (Table 9, Figure 15).

Space-time analysis with the maximum space and time windows restricted to 5% and 12%, respectively, resulted in 23 significant space-time clusters, 7 involving multiple farms and 16 involving single farms. We show the location of the 7 space-time clusters that involved multiple farms in Figure 16, and present details of the years during which each cluster occurred and the relative risk and p-value in Table 10. We have not presented details of the 16
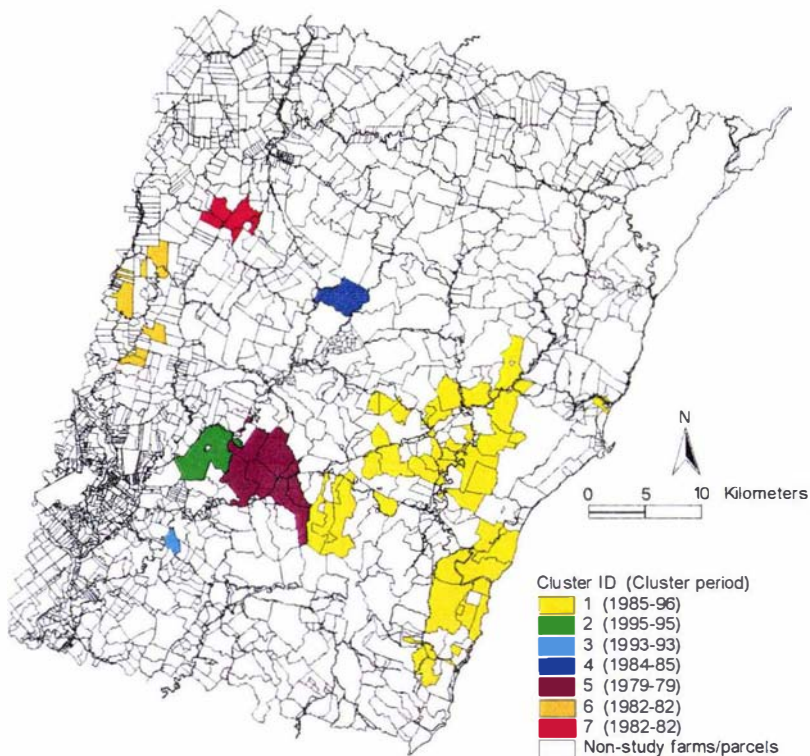
Figure 15. Location of all significant space-time clusters in the analysis with a maximum time window of 16 years (90%) and a maximum spatial window of 65 farms (50%), together with the time period during which there was significant clustering.
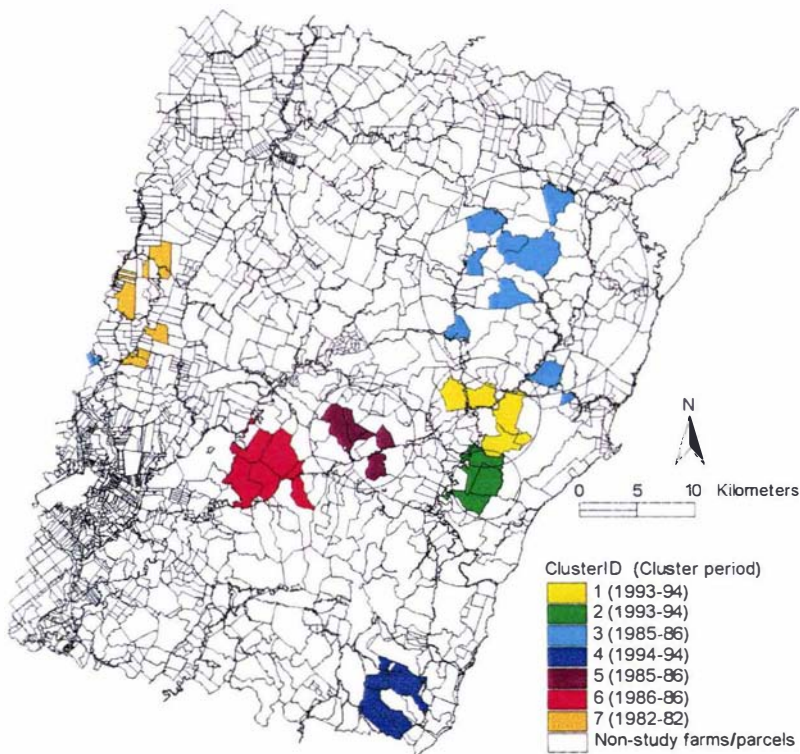
Figure 16. Location of multiple-farm space-time clusters for the analysis with a maximum time window of 2 years, together with the time period during which there was significant clustering. Clusters extending across 2 years are circled.

space-time clusters that involved single farms, as in this analysis we were interested in identifying if there was evidence of significant space-time clustering across contiguous farms within one or two-year periods, rather than identifying if the relative risk on an individual farm differed to that in the rest of the study area. Only 4 of the 8 multiple farm clusters involved clustering over a two-year period (circled in Figure 16), while the remaining 4 farms were clustered for a single year only. The three clusters in the east of the study area occurred during the period 1993-94, the three towards the middle occurred in 1985-86, while the two clusters in the west occurred in 1982.

Discussion

Calculation of the spatial and space-time scan statistics is based on point location data. In this study we represented farms as a single point by using the centroid of the largest polygon of each study farm (Figure 10). Representing areal units by a single point can give rise to distortions in spatial analyses when the areal units are not distributed randomly with respect to size (Mackereth, 1998). Farm data is often clustered by farm size, with large farms occurring together in one area and small farms in another. The density of point data representing areal units is influenced by the size and the shape of the areas that they represent, and by missing data. In our study, density was highest in a small area in the west where a number of small contiguous farms was located (encircled in Figure 10). It was lowest in areas where there was either a large number of farms not included in the data set (for example, in the south-west and north-east part of the study area), or large areas of forest (for example, in the east and south-east of the study area). Areas that included a number of larger contiguous farms had a moderate density. The application of spatial analyses that base proximity on the distance between points would not be valid in the case of this data. However, the spatial and space-time scan statistics are less sensitive to non-stationarity (uneven variation across space) of the density of points, because the analyses use a shifting circular window that varies in size from an individual point to a maximum number set by the analyst. This allows for the detection of clustering over different distances in different parts of the study area. The varying window size also makes these techniques less sensitive to local variations in density that are due to missing data points. However, they are sensitive to missing values in the attribute data used to compare the relative risk within and without each circle. We need to be cautious in interpreting the results from the spatial analyses conducted in this study given the amount of missing data. We can place reasonable confidence in the results from parts of the study area

where there are groups of contiguous study farms with few missing farms between, such as the area from the north-west through to the south-east (Figure 9). However, data are too sparse in the north-eastern and south-western parts of the study area to have confidence in the results in these areas.

Table 9: Time period, relative risk, p-value and number of farms for all significant space-time clusters with maximum space and time windows, shown in Figure 15.

Cluster ID	Cluster period	Relative risk	p-value	Number of farms in the cluster
1	1985-96	2.26	0.001	27
2	1995-95	22.22	0.001	1
3	1993-93	18.14	0.001	1
4	1984-85	4.20	0.001	1
5	1979-79	3.92	0.001	6
6	1982-82	4.77	0.003	4
7	1982-82	6.15	0.004	1

Table 10: Time period, relative risk and p-value for multiple-farm space-time clusters with maximum spatial window of 5% and time window of 12%.

ClusterID	Cluster period	Relative Risk	p-value	Number of farms in the cluster
1	1993-94	6.62	0.001	3
2	1993-94	8.62	0.001	2
3	1985-86	4.07	0.001	5
4	1994-94	5.43	0.001	3
5	1985-86	5.78	0.001	3
6	1986-86	3.81	0.001	5
7	1982-82	4.73	0.001	4

Spatial analysis of TB cattle data from beef breeding farms in an area of the Wairarapa using the spatial scan statistic showed significant clustering of both high and low rates for the five-year CI (1986-90) (Figure 12) and percent TB-positive years (Figure 14). SaTScan outputs a list of farms included in each cluster, which enabled us to map the clusters and identify their geographic location. This provides useful insights into the nature of the disease and helps

develop hypotheses regarding associated risk factors. High rates for both the five-year CI (1986-90) and percent TB-positive years were clustered in the south-eastern part of the study area, and low rates for both measures were clustered in the north-western area (Figure 12 & Figure 14). An interesting feature of the clustering patterns of these two measures was the area of overlap of the low-rate cluster for CI (1986-90) and the high-rate cluster for percent TB-positive years, and conversely of the high-rate cluster for CI and the low-rate cluster for percent TB-positive years (Figure 12 & Figure 14). The group of 13 farms included in the area of overlap had moderate values for CI (1986-90), with most falling within the range 0.5-5% (Figure 11), and moderate values for percent TB-positive years, with most falling within the range of 20-40% (Figure 13). The area of overlap of the two clusters represented a third cluster, and the combination of these clusters showed three reasonably distinct patterns of TB associated with different geographic locations in the study area. The low-low group in the north-west included 29 farms with a low percent TB-positive years and few TB-positive cattle during positive years. The low-high group in the centre included 13 farms that had a moderate percent TB-positive years and a moderate number of TB-positive cattle occurring during the positive years. The high-high group in the south-east included 39 farms with a high percent TB-positive years and a high risk of TB-positive cattle during those years.

We hypothesise that these spatial clustering patterns are related to possum habitat patterns that influence both the number of persistent possum TB clusters on a farm and the overall density of possums in the area of the farm and its surrounds. We suggest that the 39 farms in the high-high category for both percent TB-positive years and CI (1986-90) have a combination of multiple possum TB hot spots on the farm and more persistent TB infection within these hot spots over time. The multiple hot spots result in the possibility of a higher number of cattle being infected in any one year, and a higher possibility of infected possums being present in at least one of the hot spots in any one year. In a longitudinal study of an infected possum population Pfeiffer (1994) and Jackson (1995) found significant temporal variation of TB incidence in possums within individual hot spots. Given this temporal variation, a higher number of hot spots present on a farm could result in a higher chance of there being at least one infectious possum on the farm in any one year. This in turn would result in the farm experiencing a higher percent of TB-positive years compared with farms which have fewer hot spots. TB infection may be more persistent at hot spots on these farms due to a combination of habitat factors that favour the transmission of TB between possums within hot spots, and a higher density of possums in the surrounding area which provides a

large pool of susceptible possums and a pool of infected juveniles which disperse.

The space-time analysis that we conducted with the maximum size of the temporal and spatial windows was an initial exploratory analysis that showed the maximum scale of space-time clustering within the study area. The most likely space-time cluster included 27 farms located in the east of the study area during the period 1985-96 (Figure 15). This cluster included 27 of the 39 farms included in the most likely high-rate cluster for the five-year CI (1986-90). The space-time analysis uses a cumulative incidence in its calculation of relative risk within each space-time cylinder, hence the similarity of the clusters identified by the two analyses.

Space-time clustering of TB in cattle is strongly confounded by possum control operations, which are implemented on a space-time basis. Possum control operations conducted in the study area since 1990 covered an average of 30,000 hectares per operation, with intensive culling of possums in the first year of the operation and follow-up culling in subsequent years. The incidence of TB in cattle generally drops significantly in the year following a possum control operation, and remains low as long as the possum population is maintained at a low level (Anon, 1986; Pannett, 1991; Anon, 1997). The pattern of space-time clustering that we identified in the study area has been influenced by the location and timing of large-scale possum control operations. For example, the most significant space-time clustering in the east, which extended over a time period of 1985-1996, is likely to have been influenced by intensive possum control that was initially conducted in this area in 1979-80, with no further control on most farms in the cluster until 1995 (Anon 1993; Anon, 1997). The control in 1979-80 resulted in an initial reduction in the incidence of TB in cattle in this area, then the rate began to increase after 5 years which coincides with the beginning of the space-time cluster period.

By restricting the maximum time and space windows to 12% and 5% respectively we were able to investigate if there was evidence of significant space-time clustering on groups of neighbouring farms during periods of 1-2 years. There were only 7 space-time clusters that included multiple farms, and only 4 of these involved a two-year period (Figure 15), while the remaining 3 involved only one year. These results suggest that factors that influence temporal clustering of TB in possums do not commonly act across contiguous farms during the same time period. Conversely, the incidence of TB on particular farms rose and fell independently of their neighbours during one to two-year periods in most parts of the study area. Only 4 of the 7 multiple-farm clusters (encircled in Figure 16) and 4 of the 16 single-farm clusters involved a two-year period, which suggests that peaks in TB incidence mostly involved only a

single year and did not persist over consecutive years. However, these results need to be interpreted with some caution. The space-time scan statistic only identifies the most likely clusters and does not list less significant clusters that partially overlap the most likely clusters. This makes it difficult to identify any additional space-time clusters that occurred at locations overlapping the most likely cluster location, without breaking the data set into smaller time periods and conducting separate space-time analyses for each. These results are also likely to be sensitive to the number of missing data points, given the small size of clusters.

In addition to assisting with hypothesis generation the spatial and the space-time scan statistics provide very useful tools for TB surveillance purposes (Hjalmars et al., 1996). They could be used to identify individual farms or groups of farms with a higher TB incidence relative to the remainder of the area of interest in any one year, or in any group of years. This information could be used in setting priorities for the allocation of TB control resources. In this respect, these statistics could provide an objective method to define the boundaries of areas for possum control operations by identifying groups of farms with a high rate for TB incidence over the time period of interest. They could also be used to define areas that comprise low-rate clusters, where priority may be given to keeping those areas free of TB. In areas of New Zealand where possum populations are believed not to be infected with TB (referred to as Surveillance Areas), analysis for spatial and space-time clusters could be used as an early warning system to indicate that infection has entered the possum population. Any spatio-temoral clustering of TB incidence involving multiple farms in Surveillance Areas would be considered an indication that a wildlife vector, most commonly the possum, may be associated with infection, and such cases would warrant thorough investigation. We believe that given the availability of a digital map of farm locations on a regional or national basis, this analysis tool is very useful for the management of TB in New Zealand.

Bibliography

Anon. (1986). History of tb control scheme. In: (S. Hennessy (Ed)), Surveillance 13 (3): 4-8.

Anon. (1993). Wellington region operational plan for bovine tuberculosis, July 1993 - June 1998, MAF Quality Management, Masterton, New Zealand.

Anon. (1997). Vector control areas, Wellington Region (1997). MAF Quality Management, Masterton, New Zealand.

Hickling, G. (1995). Clustering of tuberculosis infection in brushtail possum populations: implications for epidemiological simulation models. In: F. Griffin, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. University of Otago Press, Dunedin, New Zealand. 174-77.

Hjalmars, U., Kulldorff, M., Gustafsson, G., and Nagarwalla, N. (1996). Childhood leukaemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. Statistics in Medicine 15 : 707-15.

Jackson, R. (1995). Transmission of tuberculosis (*Mycobacterium bovis*) by possums. Unpublished PhD thesis, Massey University, New Zealand.

Kulldorff, M., Athas, W.F., Feuer, E.J., Miller, B.A., and Key, C.R. (1998). Evaluating cluster alarms: A space-time scan statistic and brain cancer in Los Alamos, New Mexico. American Journal of Public Health 88.(9): 1377-80.

Kulldorff, M., Feuer, E.J., and Freedman, L.S. (1997). Breast cancer clusters in the Northest United States: A geographic analysis. American Journal of Epidemiology 146 (2):161-70.

Kulldorff, M. and Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. Statistics in Medicine 14 : 799-810.

Kulldorff, M., Rand, K., and Williams, KG. (1997a). Comparison with other methods. Help File. SaTScan, Version 1.0. Software for the spatial and space-time scan statistics. Bethesda, MD: National Cancer Institute. [Available at: http://dcp.nci.nih.gov/bb/SaTScan.html].

Kulldorff, M., Rand, K., and Williams, KG. (1997b). SaTScan, version 1.0. Software for the spatial and space-time scan statistics. Bethesda, MD: National Cancer Institute. [Available at: http://dcp.nci.nih.gov/bb/SaTScan.html].

Livingstone, P.G. (1997). Update on the New Zealand TB situation. In: Proceedings of a Seminar on Possum and Mustelid Control Research. National Possum Control Agencies, Wellington, New Zealand. 17-30.

Mackereth, G. (1998). Spatial data requirements for animal disease management in New Zealand. Unpublished Masters thesis, Massey University, New Zealand.

McKenzie, J.S. and Dymond, J.R. (in preparation). Automatic classification of a multispectral SPOT3 image to produce a vegetation map for modelling possum habitat in New Zealand. Geocarto International.

McKenzie, J.S., Morris, R.S., and Pfeiffer, D.U. (1997). Identification of environmental predictors of disease in a wildlife population. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 01.13.1-01.13.3.

Pannett, G. (1991). Field experience with TB control in New Zealand. In: R. Jackson (Convenor), Proceedings of a Symposium on Tuberculosis. Veterinary Continuing Education Publication No. 132, Palmerston North, New Zealand. 173-75.

Pfeiffer, D. U. (1994). The role of a wildlife reservoir in the epidemiology of bovine tuberculosis. Unpublished PhD thesis, Massey University, New Zealand.

Ryan, T.J. (1997). The New Zealand National Livestock Database. In: Epidemiology Program. Proceedings of the 10th Federation of Asian Veterinary Associations (FAVA) Congress. Cairns, Australia. 167-70.

Sanson, R.L.and Pearson, A.P. (1997). Agribase - A national spatial farm database. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.16.1 - 12.16.3.

Shortridge E.H. (1981). Tuberculosis in cattle and opossums. In: Proceedings of a Veterinary Public Health Seminar. Palmerston North, New Zealand. 83-95.

Walter, S.D. (1993). Assessing spatial patterns in disease rates. Statistics in Medicine 12 (19-20): 1885-94.

Zewdie, S. (1997). Epidemiological analysis of tuberculosis in cattle herds. Unpublished Masters thesis, Massey University, New Zealand.

100

CHAPTER 4

Automatic classification of a multispectral SPOT3 image to produce a vegetation map for modelling possum habitat in New Zealand[3]

.

Abstract

Supervised classification of a SPOT3 multispectral image provided a vegetation map to model the distribution of possums and of tuberculous possums, both at the possum home range and the farm level. The resulting raster map comprised eight vegetation classes, including three forest classes, four scrub classes and one pasture class. This is the most detailed vegetation map currently available for modelling possum habitat in New Zealand. However, the high degree of heterogeneity of species within two of the scrub classes limited the discriminatory power of models using this data.

Introduction

A national programme to control tuberculosis (TB) in cattle in New Zealand has been in place since the mid-seventies. Despite these efforts the disease is still endemic in many areas of the country and threatens our export markets (O'Neil and Pharo, 1995). The major factor hindering the eradication of TB in New Zealand is the presence of the disease in wildlife species that act as a reservoir of infection for farmed animals. The brushtail possum (*Trichosurus vulpecula*) is the major reservoir host of TB, and considerable resources are invested in the control of possums in New Zealand each year. Possum populations in many areas have been reduced to very low levels. This has resulted in a reduction in the incidence of TB in farmed cattle and deer, nevertheless continued control of possum populations is necessary in affected areas to prevent a resurgence of the possum population and an associated increase in the incidence of disease in farmed animals. Additional tools and information are required to continually refine current possum control programmes to ensure the most efficient use of resources in areas where the number of possums and of tuberculous possums have reached very low levels.

Researchers at Massey University have developed models to assist the development and evaluation of possum control strategies for TB management purposes. The TB hot spot model is a geographic model that classifies the cells of a raster map into one of three possum TB risk categories (low, moderate, high), based on a combination of vegetation and slope data (McKenzie et al., 1998). PossPOP is a geographically based simulation model of the spread of TB within a possum population that uses a vegetation map to populate the model with possums and dens (Pfeiffer, 1994). This model is used to simulate the effect of different possum control programmes at the farm level. These models are useful tools for evaluating

different possum control programmes and for differential targeting of resources based on the risk of TB in possums. They have been incorporated into a decision support system, EpiMAN(TB), to facilitate their application at the field level (McKenzie et al., 1997b). Possum control agencies have become interested in using vegetation maps to plan and implement possum control operations, as the scale and number of these operations has increased in recent years. We also required a digital vegetation map for a study to identify the association of vegetation patterns and other geographic factors with the risk of tuberculous possums being present on farms. An underlying requirement for these projects is digital vegetation data that can be used to map possum habitat. The vegetation data needs to fulfil the following criteria:

1. It can be produced in a cost-effective manner for large areas of the country.

2. It has sufficient spatial resolution to identify small patches of habitat, such as 0.25-hectare patches, shelter belts, and trees lining riverbanks.

3. It has sufficient spectral resolution to differentiate vegetation into habitat classes that represent the structural and compositional features associated with the distribution of tuberculous possums.

The distribution of TB in possums is clustered at a scale of approximately 50 meters square (Pfeiffer, 1994; Hickling, 1995) and clusters, which are commonly referred to as TB hot spots, are associated with habitat that provides highly favourable nesting sites for possums (Chapter 2). Given the association of possum TB clusters with specific habitat factors, we wanted to determine if habitat patterns at the farm level could be used to predict farms that were more likely to have TB hot spots, and hence have a higher risk of cattle becoming infected from possums. This information would be useful to predict farms that provide an environment in which TB in possums is more likely to become established, to help target disease surveillance and possum control efforts in uninfected areas surrounding known endemic areas. We were interested in investigating the heterogeneity of plant species within possum nesting habitat, plus spatial patterns of habitat as risk variables for the presence of TB hot spots on farms. We thus required vegetation data that was differentiated at a resolution of at least 50 meters square to capture habitat patterns at the scale of possum TB hot spots.

Several vegetation data sets exist in New Zealand. However, none of these met the criteria required for our purposes. Satellite imagery has been used extensively to provide vegetation data to model habitat preferences of bird and animal species (Ormsby and Lunetta, 1987;

Hodgson et al., 1988; Herr and Queen, 1993; Homer et al., 1993; and Duncan et al., 1995). In most cases Landsat MSS or Landsat TM imagery was used due to its superior spectral resolution which enables greater differentiation between plant species (Chavez and Bowell, 1988). SPOT3 multispectral imagery has good spatial resolution with 20-meter pixels, but poorer spectral resolution. As a result it has more commonly been used where spatial accuracy is important and land cover is classified into broad classes rather than highly differentiated species' classes (Ehlers et al., 1990; Millette and Edelstein, 1991; Treitz et al., 1992; Millen, 1996). More recently SPOT3 images have been used to produce vegetation maps with greater differentiation of plant species (Passini and Lacaze, 1995; Green, 1998). Automatic classification of satellite images has been shown to provide a cost-effective means of generating a current vegetation map for large areas (Dymond et al., 1996). However, there are problems associated with the production of accurate maps over large areas due to the effects of topography and varying atmospheric conditions. As a result visual interpretation of satellite imagery has been used in preference to automatic classification in the production of vegetation maps in some areas (Rasch, 1994; Rosenholm, ; Millen, 1996). We were interested in applying an automatic classification procedure in this study, as it was less time-consuming and less labour intensive than visual interpretation.

In this paper we evaluate the vegetation data sources available in New Zealand and describe the process by which we produced a vegetation map from a SPOT3 multispectral satellite image using an automated classification procedure. We discuss the results of the classification and how well the resulting vegetation map met our needs for modelling possum habitat.

Existing sources of vegetation data

Existing sources of vegetation data in New Zealand were investigated to identify the most appropriate source of vegetation data for this project. A brief description of each source is presented.

Black and white aerial photographs

A series of black and white aerial photographs was taken for the entire Wairarapa region in 1989 at a scale of 1:54000. A more recent series was collected during the period 1995-1999. Observation of black and white aerial photographs showed a significant variation in texture between different vegetation classes. For example, short, dense manuka/kanuka and gorse scrub had a smooth even appearance, whereas tall trees with large rounded crowns had a more

'lumpy' appearance. Given that the structure of the vegetation was the feature that related to the risk of TB transmission between possums, rather than the individual species *per se*, we investigated the possibility of classifying the vegetation using a texture analysis on black and white aerial photographs. However, testing of this approach showed there was too much overlap between the different classes to make this approach useful. In addition, texture information was lost in shaded areas. Given that much of the study area was hill country, this would have led to a significant loss of information.

New Zealand Land Resource Inventory (NZLRI)

The NZLRI is the principal national land resource database available in New Zealand (Eyles and Newsome, 1990; Newsome, 1991). It contains an inventory of five sets of information: rock, soil, slope, erosion and vegetation cover for the whole country at a scale of 1:63,360. It is based on map units of variable size that represent homogeneous land units derived from a combination of rock type, soil type, and slope. The minimum map unit area is approximately 60 hectares. Vegetation in each unit is recorded as either covering more than or less than forty percent of the unit. Vegetation covering less than 10 percent is not recorded. Thus, combinations of vegetation types are commonly recorded and there is no information on the distribution or location of vegetation types within each unit (Blaschke et al., 1981). The NZLRI was surveyed between 1973 and 1979, hence some of the vegetation information is now out-of-date, particularly in areas where exotic forests have been planted and scrub has either been cleared or regenerated (Wilde, 1996).

Whilst the NZLRI has a detailed vegetation legend, it was not appropriate for use in this project as the distribution and location of vegetation categories were not sufficiently specific. For example, a vegetation classification of the Gisborne district generated from a Landsat TM image in 1994 showed the total area of manuka/kanuka cover to be 40,000 hectares (Dymond et al., 1996). For the same area the NZLRI showed the total area covered by manuka/kanuka to be much larger because the vegetation classification was generalised across the entire map unit within which it occurred.

The vegetative cover map of New Zealand

The vegetative cover map of New Zealand was compiled between 1981 and 1986 and published at a scale of 1:1 000 000, (Newsome, 1987; Newsome, 1991). This database classifies vegetation into 49 classes with a legend similar to that of the NZLRI, from which it

was derived. Although this database does have more complex vegetation associations, including mixes of forest, scrub and pasture, it did not show enough scrub species to be used as the primary source of vegetation information for this project. As was the case with the NZLRI, this database also lacks detailed location and distribution information for vegetation classes.

Ecological survey of New Zealand indigenous forests

All primary indigenous forests were mapped by the New Zealand Forest Service in the early 1970s as a part of an ecological survey of New Zealand indigenous forests (Nichols, 1974). The classification of forest classes is more detailed than the NZLRI and includes 18 classes. Although the mapping was done over 20 years ago, the primary indigenous forest has changed little in that time. These maps do not contain any information on scrub or small remnants of indigenous forest outside the major forest areas, and are not in digital form.

The New Zealand Department of Conservation in Wellington does have a digital map of indigenous forest in the Tararua Forest Park, which forms the western border of the study area. However, this map has no detailed classification of vegetation cover on farmland outside the forest park boundary.

Protected Natural Areas Programme

The Department of Conservation is implementing a Protected Natural Areas Programme (PNAP) designed to establish a network of reserves which is representative of the full range of New Zealand's natural diversity as embodied in the Reserves Act 1977 (Leathwick et al., 1994). It involves the assessment of remnants of natural vegetation to allow identification of those with high conservation values located outside the existing reserves' network which should be given some form of protection. The eastern Wairarapa area, which includes part of the study area, was surveyed during 1998 (Anon, 1997). The outline of each remnant was drawn onto a 1:50000 topographic map sheet and the species present within each remnant were recorded. These are divided into dominant species and other species, but the relative proportions of each are not recorded. The species in each remnant are listed without being keyed to a vegetation legend. The data for this area was not available in digital form at the time of undertaking this study.

New Zealand Land Cover Data Base

Terralink, the commercial arm of the New Zealand government's land information

department, began developing a national land cover database from current satellite imagery in 1996 (Millen, 1996). The land is classified into classes within artificial, cultural and natural landscape categories. Natural landscapes include grassland, scrub, indigenous forest, exotic forest, bare ground, inland wetlands, coastal wetlands, inland water, and mangrove. The database comprises vector maps that are generated by visual interpretation of SPOT3 multispectral images, with a minimum mapping unit of 1 hectare. As of June 1998, Terralink had completed the land cover database for all of the North Island and approximately half of the South Island. The major use of this database to date has been the identification of areas planted in exotic forest and areas of indigenous forest.

While these maps have a high degree of spatial accuracy, considerable vegetation cover information has been lost by the broad categories used for classification and by using a minimum mapping unit of 1 hectare. Patches of vegetation less than one hectare have been incorporated into the surrounding vegetation class, resulting in the loss of habitat such as rows of willows along riverbanks. Heterogeneity within a vegetation patch is not represented as the patch is classified according to the dominant vegetation class present. This database was not sufficiently detailed to model possum habitat with sufficient accuracy for our purposes.

Landcare Research has produced similar data for the Bay of Plenty region and for the Gisborne and Tasman districts, using SPOT3 multispectral images for the former and Landsat TM for the latter.

Satellite imagery

Landcare Research was the sole distributor of satellite images in New Zealand until 1996. They hold an archive of all satellite images (Landsat, SPOT and NOAA) that have been purchased by New Zealand organisations until 1996, when Terralink also began to acquire SPOT3 satellite imagery. While Landsat TM images have high spectral resolution due to the collection of data in 11 spectral bands, we chose not to use this imagery because of its coarser spatial resolution at 30 square meters, which resulted in some smaller patches of habitat being missed. SPOT3 multispectral images have a better spatial resolution at 20-meter pixels. This enables smaller patches of land cover to be identified, down to individual rows of trees. While the spectral resolution is poorer than Landsat imagery, it provides sufficient detail to differentiate broad vegetation classes such as pasture, scrub, indigenous and exotic forest. Given the better spatial resolution and the adequate spectral resolution, we chose to use a SPOT3 multispectral image as the source of vegetation data for this project. Historical SPOT

multispectral images for the study area were held in the archives. We used these to gain an impression of the land cover classes that could be differentiated using this imagery, but purchased a current image for the project.

Method

Study area

The study area was a 60 kilometer square area in the north-eastern part of the Wellington Region, known as the Wairarapa (Figure 17). The area corresponded with the coverage of a SPOT3 satellite image that was used for this study. This area is a pastoral farming area dominated by extensive sheep and beef farming, with a topography of rolling to steep hills. In the western part of the study area vegetation is predominantly improved pasture species, with patches of manuka/kanuka (*Leptospermum scoparium/Kunzea ericoides*) in gullies, some pine (*Pinus radiata*) plantations and podocarp forest remnants. There are some plantings of exotic species such as willows (*Salix* species) and poplars (*Populus* species) for soil conservation purposes, with willows dominant along river banks. The eastern part of the study area is dominated by a range of hills that runs parallel to the coast and is covered predominantly with pine forest. There are also large areas of beech (*Nothofagus* species) or podocarp/broadleaved forest on the hills, and mixed patches of gorse (*Ulex europeaus*) and manuka (*Leptospermum scoparium*) occur where the forest has been cleared.

Image analysis

An image of the study area was acquired on 3 March 1994 (Figure 17). It had dense cloud in the north-western area and a few small patches of cloud elsewhere (Figure 19), but was generally of good quality with an off-nadir view angle of 1.8°. The image was ortho-rectified to the New Zealand Map Grid (NZMG), using a 20-meter raster digital elevation model. Even though the image had a low look angle of 1.8°, we had it orthorectified as much of the terrain covered by the image was steep hill country. Burgess and McNeill (1992) calculated that at a height of 300 meters this angle would cause a planimetric distortion of approximately 20 meters. We needed to compare the area of each habitat type on farms throughout the study area, thus it was important to remove the distortion that steep hills would have introduced to the area calculations had the image not been orthorectified.
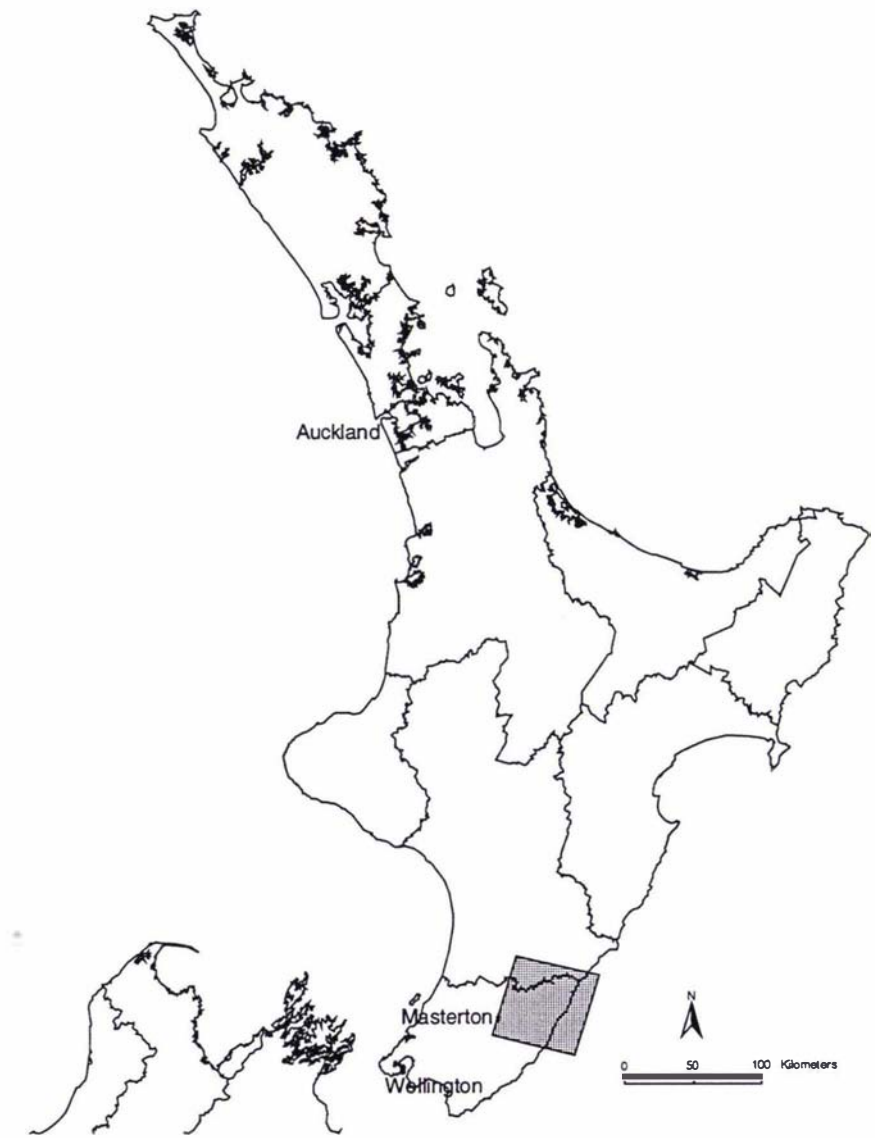
Figure 17. Regional map of the North Island of New Zealand showing the location of the study area in the northern part of the Wellington Region.

We conducted the image analysis in ERDAS Imagine (ERDAS Inc, Atlanta, Georgia, USA) using a supervised classification approach. Our initial aim was to identify as many classes of habitat as possible. We initially listed 13 habitat classes that were associated with the distribution of tuberculous possums, and which we thought were likely to be distinguishable on a SPOT3 multispectral image. These were pasture, gorse *(Ulex europaeus)*,

manuka/kanuka (*Leptospermum scoparium/Kunzea ericoides*), willows (*Salix* species), pine forest (*Pinus radiata*), beech (*Nothofagus* species), groves of tree ferns (*Cyathea* or *Dicksonia* species), podocarp-broadleaved forest, broadleaved scrub, swamps, shelter belts of large mature trees (for example, macrocarpa (*Cupressus macrocarpa*), poplar (*Populus)* and pine *(Pinus radiata)*), and mahoe (*Melicytus ramiflorus*) stands.

We displayed the image by assigning the near infrared, red and green spectral bands to the red, green and blue display guns, respectively. We used field knowledge, colour photographs and black and white aerial photographs to identify the colours and tones associated with the habitat classes listed above. We then manually drew over several training plots for each of the classes that could be distinguished and applied a maximum likelihood classifier to create an initial vegetation classification.

Due to budget and time constraints we were not able to undertake extensive field checking of the vegetation map. We thus used a combination of methods to check the accuracy of the classification. We checked about 50 accessible sites in the field. In addition to this, we used existing data from the Department of Conservation's PNA surveys to check other areas of the map. We checked a total of 127 PNA sites against the vegetation map in the four areas shown in Figure 18. As the description of PNA sites only divided the species into 'dominant' and 'other species' with no information on the proportion of the site covered by each species, we could only base our accuracy assessment on the presence or absence of species rather than on the area covered by each species.

We manually corrected obvious errors in the spectral classification using raster editing software. Many areas of pasture were classified as bare ground as the image was acquired at the end of the summer when pasture on the hills was very dry. We recoded bare ground across the whole map as pasture, which resulted in rock and sand along the coast also being recoded as pasture. This was acceptable as we were only interested in using the data on farmland. We did not correct areas where pine had been misclassified as beech as these two classes were similar in terms of their risk of TB in possums, and the misclassified pixels generally occurred around the perimeter of pine forest patches. We recoded crops and high quality pasture that had been misclassified as Podocarp/broadleaf, to Pasture. We edited small areas covered by cloud with the use of aerial photographs, and left large areas of cloud unclassified.

We did not smooth the data because we did not wish to eliminate small patches of habitat, and we wanted to retain the full extent of habitat heterogeneity.
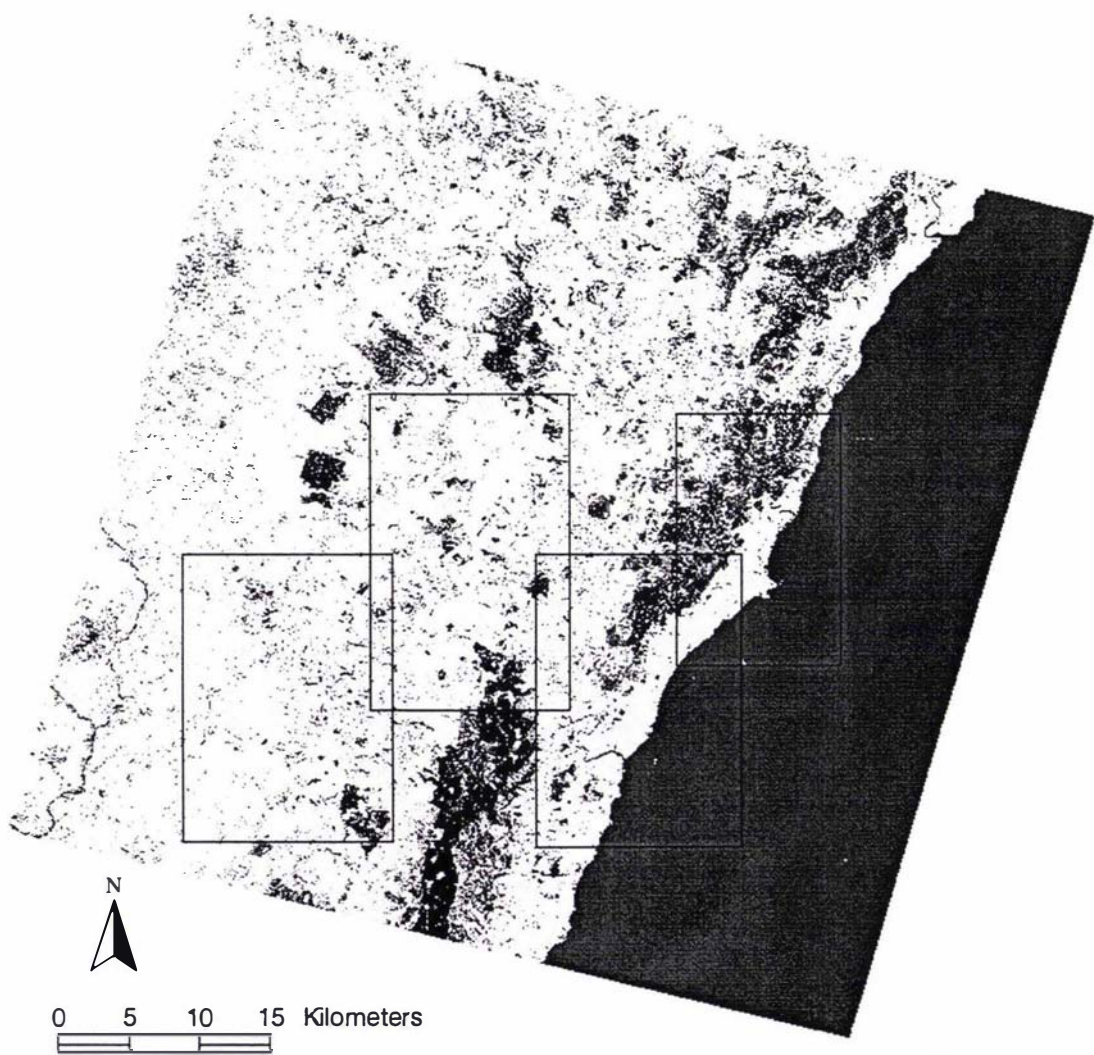
Figure 18. Map showing the location of the PNA survey sites used for checking the vegetation classification.

Results

The SPOT3 image did not have a spectral range that was adequate to differentiate all the vegetation classes that we listed above. Consequently, we had to revise the list to a smaller number of broader vegetation classes. We show the final map legend of the vegetation classes that we could distinguish on the basis of spectral signature in Table 11. We show the results of the accuracy checking in Table 12. The figures in Table 12 represent the number of PNA checking sites at which each vegetation class was accurately classified on the vegetation map. We have used asterisks to show the classes into which groups of plant species were misclassified, as we did not have sufficient field data to quantify the relative distribution of

the misclassified classes. Overall, vegetation classes were accurately classified at 90% of the 127 sites used for checking, with a range of 73-100%. Classification results for each class are discussed in more detail below.

Pine

Pine was correctly classified on the vegetation map as being present at only 30 of the 41 (73%) PNA sites at which it was recorded. Errors of omission were predominantly due to sparse pine trees on pasture being misclassified as beech. Other species were misclassified on the vegetation map as Pine at 17 sites where pine had not been recorded in the PNA survey. Field checking showed that this was mostly due to podocarp species and small patches of beech being misclassified as pine.

Beech

Beech was correctly classified on the vegetation map at 13 of the 16 (81%) PNA sites at which it was recorded. Field checking showed that stands of beech trees covering less than 1 hectare (approximately) were misclassified as pine. Where beech occurred as the dominant species over a large area, the accuracy of classification was high. Other species were misclassified as beech at 61 sites where beech was not recorded as being present in the PNA survey. Field checking showed that the most common reason for this was mixed pixels of pine or podocarp species and pasture being misclassified as Beech.

Podocarp/broadleaf

Podocarp/broadleaved species were correctly classified on the vegetation map at 81 of the 93 (87%) PNA sites at which they were recorded. Errors of omission were predominantly due to mixed pixels of low-density podocarp trees on pasture being misclassified as beech. Field checking showed that podocarp and broadleaf forest species were classified with a high degree of accuracy. However, the accuracy for broadleaved scrub species such as mahoe (*Melicytus ramiflorus*), five finger (*Pseudopanax arboreus*), *Coprosma* species and lace bark (*Hoheria populnea*) was lower, particularly if these species were mixed with manuka/kanuka. Broadleaved scrub species were classified as either podocarp-broadleaf, manuka/gorse or manuka/kanuka under different circumstances.

Other species were misclassified into the Podocarp/broadleaved class at 9 sites where podocarp/broadleaved species were not recorded as being present in the PNA survey. Field checking showed that this was predominantly due to young vigourously growing gorse and
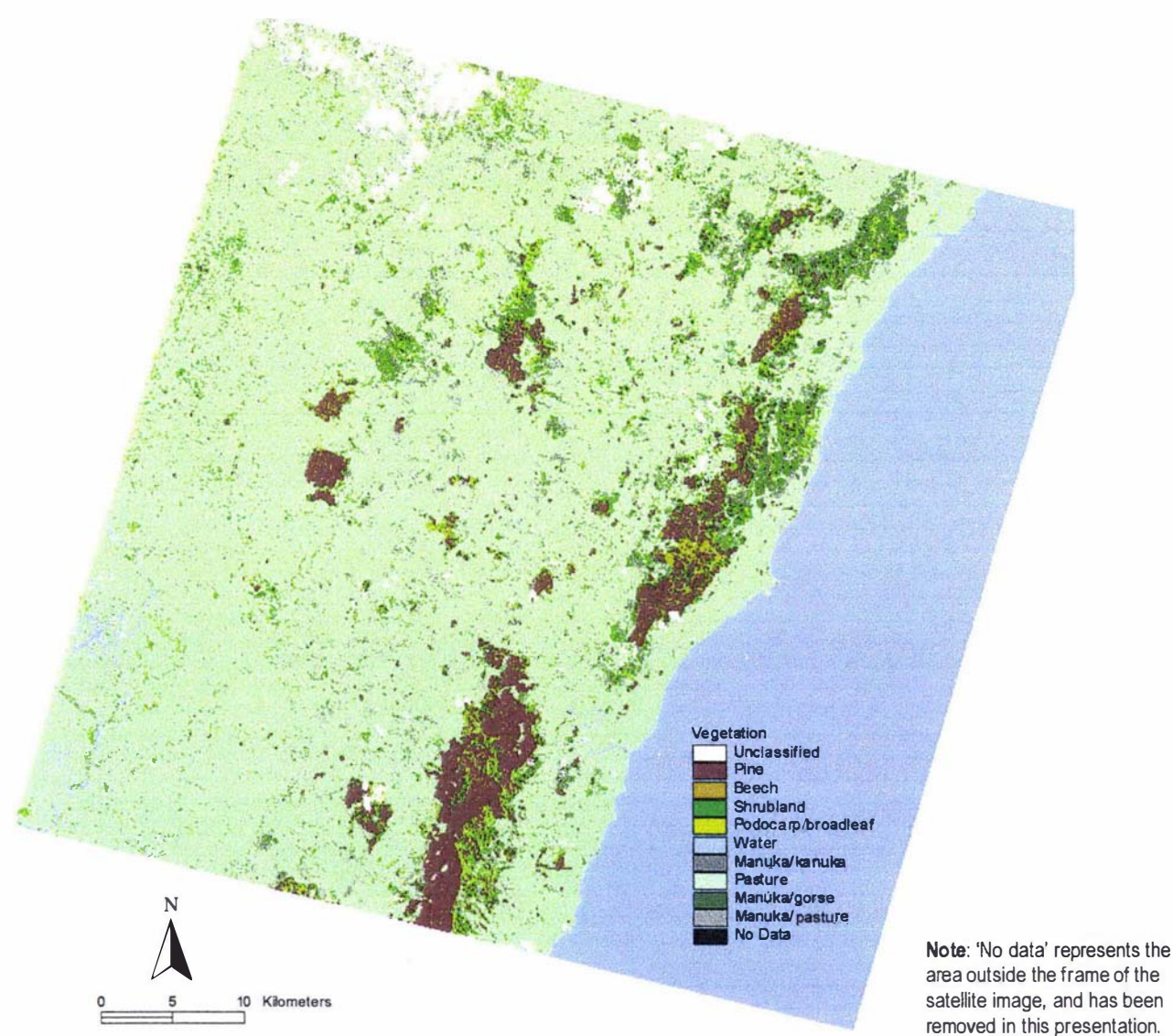
Figure 19. Vegetation map produced by automatic classification of a SPOT3 multispectral image.

vigourously growing crops being misclassified as Podocarp/broadleaf. This class also included dense patches of willows, kowhai and tree ferns, plus some exotic species such as gums, poplars, and other deciduous garden trees.

Manuka/kanuka and Manuka/pasture

These two classes were combined for error checking as they both included manuka/kanuka trees. The difference between the two classes was the density of trees, with Manuka/kanuka having a higher density than Manuka/pasture. We could not check the accuracy with which

Table 11. Description of the classes in a vegetation map that was produced by automatic classification of a SPOT3 multispectral image

Vegetation class	Description
Pine	*Pinus radiata* of all ages, except very young
Beech	*Nothofagus* species
Podocarp/broadleaf	Podocarp species and some broadleaved forest and scrub species. Also included Eucalypt species, poplars (*Populus* species), exotic garden species and dense stands of tree ferns (*Dicksonia* and *Cyathea* species), willow (*Salix* species) and kowhai (*Sophora* species)
Manuka/kanuka	Pure stands of *Leptospermum scoparium* and *Kunzea ericoides* (higher than 2 meters)
Manuka/pasture	Low-density stands of *L. scoparium* and *K. ericoides* on pasture
Shrubland	Kowhai (*Sophora* species), willow (*Salix* species), low-density gorse (*Ulex europaeus*), scattered manuka/kanuka
Manuka/gorse	Short dense manuka (*Leptospermum scoparium*), mingimingi (*Leucopogon fasciculatus*) and gorse (*Ulex europaeus*) in pure or mixed patches
Pasture	High quality and rough pasture comprising native and introduced grass species
Water	Rivers, lakes and sea
Unclassified	Residential areas, cloud cover and steep shaded faces

sites had been classified into each of these classes separately as we had no information on the density of plants in the PNA survey data. A total of 123 of the 127 PNA sites was recorded as having either manuka or kanuka trees present, and all sites were accurately classified on the vegetation map. Three sites were classified as Manuka/kanuka on the vegetation map when these species were not reported as present in the PNA survey. Given the list of species recorded as being present at these sites, it was most likely that broadleaved scrub had been classified as Manuka/kanuka. Field checking showed that the Manuka/kanuka class generally represented pure stands of taller (over 2 meters) manuka and/or kanuka trees, although in some cases it did include shorter dense patches of these species.

Table 12. A cross-classification matrix showing the results of accuracy checking the classification of the seven non-pasture vegetation classes against data collected during a PNA survey at 127 locations

PNA survey (true data)	Number of accurately classified sites on the vegetation map (interpreted data)						Num. PNA sites where present[1]	Percent accuracy
	Pine	Beech	Pod/broad-leaf	Man/kan Man/pas	Shrub-land	Manuka/gorse		
Pine	30	*					41	73
Beech	*	13					16	81
Pod/broadleaf	*	*	81	*	*	*	93	87
Manuka/kanuka				123	*		123	100
Shrubland					124		124	100
Manuka/gorse			*			60	60	100

* Indicates the vegetation map classes into which each species group was misclassified.

[1] A total of 127 sites was used for checking with data from the PNA survey. There were multiple vegetation classes present at some of the sites used for checking, hence the numbers in this column total to more than 127.

Shrubland

This was a heterogeneous category that represented mixed pixels of several scrub species at a low density on pasture. A total of 124 of the 127 PNA sites was recorded as containing at least one of the species included in this class, and all were accurately classified on the vegetation map as Shrubland. Field checking showed that each of the species included in this class was also represented in one of the other classes, when it was present at a higher density (for example, kowhai, willows and manuka/kanuka) or was at a more vigorous stage of growth (for example, gorse).

Manuka/gorse

This was also a relatively heterogeneous class that included a range of scrub species growing at a higher density than that represented by Shrubland. A total of 60 of the 127 PNA sites was recorded as containing at least one of the species included in this class and all were accurately classified on the vegetation map as Manuka/gorse.

A copy of the final vegetation map is shown in Figure 19. The forest-covered hills towards the eastern coast are clearly delineated, as is the pattern of increasing heterogeneity of vegetation cover as one moves eastward towards these hills. This heterogeneity is associated

with poorer quality hill country where there are large patches of gorse and manuka scrub that are interspersed with patches of exotic and indigenous forest.

Discussion

The most accurate classification results from SPOT3 multispectral images have been obtained by limiting classes to general land cover types and broad vegetation classes such as pasture, shrubland, indigenous and exotic forests (Ehlers et al., 1990; Millette and Edelstein, 1991; Treitz et al., 1992; Millen, 1996). However, we were interested in differentiating plant species as much as possible, in particular the scrub species as these are an important source of possum habitat in the Wairarapa, particularly in the more extensive hill country towards the coast. There is a considerable range in the possum densities supported by different scrub species (Batcheler and Cowan, 1988) and likewise a difference in the risk of possum TB hot spots. The density of scrub and tree species also influences the density of possums. It was thus important to differentiate both the scrub species and the density of vegetation as much as possible, to model accurately the distribution of possums and of tuberculous possums.

We produced a map with eight vegetation classes plus bare ground and water classes using an automated supervised classification of a SPOT3 multispectral image. The classification of land cover into the different vegetation classes was influenced by a combination of plant species and structural features of the plants, such as height and density. The eight vegetation classes included three tree classes, four scrub classes and a pasture class (see Table 11). Species were accurately classified into the appropriate vegetation class at an average of 90% of the 127 sites used for accuracy checking (Table 12). A point worth noting is that the use of PNA sites for accuracy checking resulted in an underestimation of the accuracy with which pine species are classified on SPOT3 multispectral images. The low accuracy rate resulted from there being too few pine trees at many of the PNA sites, where they were recorded as present, to be classified as Pine on the vegetation map. This occurred because the focus of the PNA survey was on indigenous forest, hence pine was not a dominant species at most of the 41 PNA sites where it was recorded. Field checking showed that pine was classified with a high degree of accuracy except where it was present in a mixed pixel with pasture, such as on the perimeter of pine forest stands. One of the major applications of SPOT3 multispectral imagery in New Zealand has in fact been to identify pine forest plantations. The high degree of accuracy recorded for the combined Manuka/kanuka and Manuka/pasture classes, plus Shrubland and Manuka/gorse relates both to the heterogeneity of plant species included in the

latter two classes and the fact that manuka/kanuka was represented in all four classes. If manuka/kanuka was recorded as being present at a PNA site, the classification on the vegetation map was considered correct for any one of these four classes. The classification of manuka/kanuka into each of these classes depended on a combination of the height and density of manuka/kanuka plants and the other species with which it was mixed, and we could not determine this level of detail from the PNA survey results.

The three tree classes were Pine, Podocarp/broadleaf, and Beech (Table 11). Pine was a very homogeneous class that almost exclusively included *P. radiata*, plus small patches of beech and occasional pixels containing podocarp species. Podocarp/broadleaf was a more heterogeneous class that included podocarp and broadleaf forest species, plus some broadleaved scrub. It also included dense patches of willows, kowhai and tree ferns, some exotic tree species such as gum (*Eucalypt* species) and ornamental garden trees, plus patches of very vigorously growing gorse and crops. Beech included areas of beech forest greater than 1 hectare (approximately), plus mixed pixels of tall trees (predominantly pine and podocarp) and pasture. Almost all areas of pine forest and podocarp forest had a strip of beech classified around their perimeter where mixed pixels of pine or podocarp trees and pasture occurred. Single trees on pasture were also classified as beech.

The four scrub classes varied considerably in the heterogeneity of species that they included. Manuka/pasture and Manuka/kanuka were relatively homogeneous classes that included pure stands of manuka and/or kanuka trees that were taller than 2 meters (approximately), with the former representing lower density stands on pasture and the latter higher density stands. The Manuka/kanuka class also included a small proportion of manuka/kanuka bushes that were less than 2 meters tall, manuka mixed with gorse, and manuka/kanuka mixed with broadleaved scrub species. The Manuka/gorse and Shrubland classes were the most heterogeneous of the eight vegetation classes, with the former capturing high-density areas of scrub species and the latter low-density areas. Manuka/gorse included both pure and mixed patches of dense gorse, short dense manuka/kanuka,and/or *Leucopogon* or *Cassinia* species. It also included some patches of broadleaved scrub, particularly when these were mixed with manuka/kanuka trees. Shrubland included low-density gorse, rows of kowhai and/or willows lining riverbanks, very scattered manuka/kanuka trees, and possibly very low-density areas of broadleaved scrub species.

Three different pasture classes were identified on the image, ranging from vigorously growing high quality pasture to poor quality rough pasture. These were combined into one pasture

class as pasture quality was not significant in terms of possum habitat. There was some overlap between rough pasture and very low-density scrub such as gorse scattered on pasture.

The vegetation map that we produced in this study was more useful than existing data sets of New Zealand vegetation for modelling possum habitat because the spatial distribution of vegetation classes was accurately represented, and because of the finer discrimination of plant species. This data represented well the heterogeneity of vegetation classes within areas of habitat and enabled us to identify specific habitat patterns associated with possum TB risk on farms. The major limitation of this data set was the diversity of plant species that were included in some of the vegetation classes, in particular the scrub classes. Shrubland and Manuka/gorse were the most diverse classes in terms of the floristic mix found in different localities for the same classified vegetation type. As a result they probably represented different TB risks at different locations despite being classified as the same vegetation class. The Shrubland class in particular contained species which supported extremely different densities of possums, and which may have had very different risks in terms of TB possums associated with them. For example, willows can support high-density possum populations where the trees are large and have developed holes in their branches and trunks, in which possums nest (Brockie, 1991). Such trees also provide an environment that favours the spread of TB between possums and thus constitutes a high risk for the presence of possum TB hot spots (McKenzie et al., 1997a). Very scattered manuka/kanuka on pasture supports very low possum densities and is not associated with a significant risk of TB infection in possums. Low-density gorse falls between scattered manuka/kanuka and willows in terms of the possum population it supports. When using this vegetation data to model the distribution of possums and/or the risk of TB hot spots all areas of habitat that comprised these plant species had the same possum density and TB hot spot risk attributed to them, as they had been classified into the same vegetation class (Shrubland). The Manuka/gorse class contained species that support similar possum densities, but which may have different TB risks associated with them. This generalisation of the vegetation data within some of the vegetation classes reduced the specificity of the prediction of possum TB risk for patches of habitat when using this data. Greater differentiation of the scrub species would significantly improve the precision of the models that use digital vegetation data to predict the distribution of both possums and TB possums at the habitat patch level and the farm level.

The use of vegetation maps will become increasingly important in the control of TB in New Zealand as the number of possums in infected areas is reduced, and control measures will

need to be differentially targeted at higher risk areas to ensure the efficiency of these programmes is maintained. The technology available for both the capture of spatial data and the analysis of this data is currently developing rapidly. The remote sensing industry is expanding and alternative sources of satellite imagery with higher spectral and spatial resolution have become available since this project was undertaken. SPOT 4, which has a mid-infrared band with 20-meter resolution is now operational. The mid-infrared band will enable greater differentiation of plant species from these images compared with earlier SPOT images (Wilde, 1994; Chavez and Bowell, 1988). New satellites are being built to produce images with even greater spatial and spectral resolution than that described above. Geographic information systems are becoming easier to use, enabling a wider range of people to use the cartographic and analytic functions of this software. Decision support systems that use geographic data are also becoming available (McKenzie et al., 1997). It is thus important to invest further effort into identifying a source of digital vegetation data that will enable TB control managers and researchers to take full advantage of this evolving area of technology.

Acknowledgements

Bibliography

Anon. (1997). Eastern Wairarapa Ecological District. Protected Natural Areas Programme Reconnaissance Survey, Wellington Conservacncy, Department of Conservation, Wellington, New Zealand.

Batcheler, C.L. and Cowan, P.E. (1988). Review of the status of the possum (*Trichosurus vulpecula*) in New Zealand, for the Agricultural Pest Destruction Council. Ministry of Agriculture and Fisheries, Wellington, New Zealand.

Blaschke, P.M., Hunter, G.G., Eyles, G.O., and van Berkel, P.R. (1981). Analysis of New Zealand's vegetation cover using land resource inventory data. New Zealand Journal of Ecology 4 : 1-19.

Brockie, R.E. (1991). Ecology of an uninfected farm possum population. In: Dr R. Jackson (Convenor), Symposium on Tuberculosis. Foundation for Continuing Education of the N.Z.V.A., No.132, Palmerston North, New Zealand. 53-66.

Burgess, D.W.and McNeill, S.J. (1992). Terrain distortion in satellite imagery. In: Proceedings of the 6th Australasian Remote Sensing Conference. Wellington, New Zealand.

Chavez, P.S. and Bowell, J.A. (1988). Comparison of the spectral information content of Landsat thematic mapper and SPOT for three different sites in the Phoenix, Arizona region. Photogrammetric Engineering and Remote Sensing 54 (12): 1699-708.

Duncan, B.W., Breininger, P.A., Schmalzer, P.A., and Larsen, V.L. (1995). Validating a Florida Jay habitat suitability model, using demography data on Kennedy Space Center. Photogrammetric Engineering and Remote Sensing 61 (11): 1361-70.

Dymond, J.R., Page, M.J., and Brown, L.J. (1996). Large area vegetation mapping in the Gisborne district, New Zealand, from Landsat TM. International Journal of Remote Sensing 17 (2): 263-75.

Ehlers, M., Jadkowski, M.A., Howard, R.Ri., and Brostuen, D.E. (1990). Application of SPOT data for regional growth analysis and local planning. Photogrammetric Engineering and Remote Sensing 56 (2): 175-80.

Eyles, G.O. and Newsome, P.F. (1990). A decade of experience using a New Zealand geographic information system. New Zealand Geographer 46 (1): 43-47.

Green, M. (1998). Research: the basis of hazard policy and its implementation. Research into wildfire hazard at the Wellington Regional Council. In: D. M. Johnston, and P. A. Kingsbury (Comps), Proceedings of the Natural Hazards Management Workshop. Information Series, 45. Institute of Geological and Nuclear Sciences Limited and Canterbury Regional Council, Lower Hutt, New Zealand.

Herr, A.M. and Queen, L.P. (1993). Crane habitat evaluation using GIS and remote sensing. Photogrammetric Engineering and Remote Sensing 59 (10): 1531-38.

Hickling, G. (1995). Clustering of tuberculosis infection in brushtail possum populations: implications for epidemiological simulation models. In: F. Griffin, and G. de Lisle, Tuberculosis in Wildlife and Domestic Animals. University of Otago Press, Dunedin, New Zealand. 174-77.

Hodgson, M.E., Jensen, J.R., and Mackey, Jr.H.E. (1988). Monitoring wood stork foraging habitat using remote sensing and geographic information systems. Photogrammetric Engineering and Remote Sensing 54 (11): 1601-7.

Homer, C.G., Edwards, Jr.T.C., Ramsey, R.D., and Price, K.P. (1993). Use of remote sensing methods in modelling sage grouse winter habitat. Journal of Wildlife Management 57 (1): 78-84.

Leathwick, J., Clarkson, B., and Ward, C. (1994). Use of GIS for Protected Natural Area Programme Surveys: Waiapu Ecological District, a Case Study. In: P. R. Stephens (Ed), Proceedings of the 4th National Remote Sensing and Geographic Information Systems Workshop. Manaaki Whenua - Landcare Research New Zealand Ltd, Aokautere, New Zealand. 69-74.

McKenzie, J.S., Morris, R.S., Tutty, C.J., and Pfeiffer, D.U. (1997b). EpiMAN-TB, a decision support system using spatial information for the management of tuberculosis in cattle and deer in New Zealand. In: R. T. Pascoe (Ed), Proceedings of the Second Annual Conference on Geocomputation. Spatial Information Research Centre, University of Otago, Dunedin, New Zealand. 271-77.

McKenzie, J.S., Pfeiffer, D.U., and Morris, R.S. (1998). Geographic Modelling of the Distribution of Tuberculosis in Possums in New Zealand. In: P. Firns (Ed), Proceedings of the Third Annual Conference on Geocomputation. Spatial Information Research Centre, University of Otago, Dunedin, New Zealand. 209-217.

Millen, P. (1996). New Zealand land cover database. In: M. F. White, and P. R. Stephens (Eds), Proceedings of the 5th National Remote Sensing and Geographic Information Systems Workshop. Landcare Research New Zealand Ltd, Palmerston North, New Zealand. 27-28.

Millette, T.L.and Edelstein, E.L. (1991). Regional data layer development for the Vermont GIS from image processing of SPOT imagery. In: ACSM-ASPRS Annual Convention. Technical Papers (GIS), Volume 4.

Newsome, P.F.J. (1987). The vegetative cover of New Zealand. Water and Soil Miscellaneous Publication No. 112, Ministry of Works and Development, Wellington, New Zealand.

Newsome, P.F.J. (1991). The National GIS Database. In: P. R. Stephens, and M. F. White (Eds), Proceedings of a Workshop on Remote Sensing and Geographic Information Systems for Natural Resource Management. DSIR Land Resources, Technical Record 59, Lower Hutt, New Zealand. 52-53.

Nichols, J.L. (1974). Forest Service mapping series 6, Sheet No. 14, Tararua. New Zealand Forest Service, Rotorua, New Zealand.

O'Neil, B.D. and Pharo, H.J. (1995). The control of bovine tuberculosis in New Zealand. New Zealand Veterinary Journal 43 (7): 249-55.

Ormsby, J.P. and Lunetta, R.S. (1987). Whitetail deer food availability maps from thematic mapper data. Photogrammetric Engineering and Remote Sensing 53 (8): 1081-85.

Passini, M.-F. and Lacaze, B. (1995). Vegetation mapping in Sierra Juarez (Baja California, Mexico) from SPOT data. Geocarto International 10 (2): 31-37.

Pfeiffer, D. U. (1994). The role of a wildlife reservoir in the epidemiology of bovine tuberculosis. Unpublished PhD thesis, Massey University, New Zealand.

Rasch, H. (1994). Mapping of vegetation, land cover, and land use by satellite - experience and conclusions for future project applications. Photogrammetric Engineering and Remote Sensing 60 (3): 265-71.

Rosenholm, D.G. Land use and vegetation mapping by satellite - some experiences gained by SSC Satellitbild in 1987-1993. ITC Journal.

Treitz, P.M., Howarth, P.J., and Gong Peng. (1992). Application of satellite and GIS technologies for land-cover and land-use mapping at the rural-urban fringe: a case study. Photogrammetric Engineering and Remote Sensing 58 (4): 439-48.

Wilde, R.H. (1994). Landsat TM data to upgrade the vegetation layer of the NZLRI national GIS. In: P. R. Stephens (Ed), Proceedings of the 4th National Remote Sensing and Geographic Information Systems Workshop. Manaaki Whenua - Landcare Research New Zealand, 75-83.

Wilde, R.H. (1996). Updating the vegetation layer of the NZLRI national database using satellite imagery. New Zealand Geographer 52 (1): 12-20.

CHAPTER 5

Application of GIS and spatial analytical techniques to statistical modelling of possum-associated TB risk on farms

Introduction

The brushtail possum (*Trichosurus vulpecula*) acts as a reservoir host of *Mycobacterium bovis* in New Zealand and is the major source of tuberculosis (TB) infection for farmed cattle and deer (Jackson, 1995; Morris and Pfeiffer, 1995; Livingstone, 1997). Infected possum populations are estimated to cover approximately 24% of the land area of New Zealand in areas referred to as Vector Risk Areas (VRAs) (Livingstone, 1997). Both the size and number of VRAs have increased steadily since TB was first identified in possums in the late 1960s (Jackson, 1995). A key objective of the Animal Health Board, the farmer representative organisation responsible for making policy with respect to the TB management programme, is to restrict the increase in size of existing VRAs (Animal Health Board, 1995). The major method that is used to restrict this growth in size of VRAs is the generation of buffer zones with very low-density possum populations in areas that are adjacent to known infected areas. Despite this strategy some VRAs have gradually increased in size over time. Additional information to predict farms at greatest risk of developing a possum TB problem in areas surrounding VRAs, would help target possum control and TB surveillance efforts. Such information may make the implementation of buffer zones more effective.

While the prevalence of TB in possums on infected farms is generally low, in the order of 1-2% (Anon, 1986; Hickling, 1991), the disease in possums is clustered within small areas within farms (Hickling, 1989; Pfeiffer, 1994) where the prevalence can reach high levels (Pfeiffer, 1994; Hickling, 1995; Jackson, 1995). The size of possum TB clusters, commonly referred to as hot spots, has been shown to be small with a cross-sectional width in the order of 20-40 meters (Hickling, 1991; Pfeiffer, 1994; Pfeiffer et al., 1995). These clusters are associated with habitat factors that favour the transmission of TB between possums in and around their dens (see Chapter 2). Cattle become infected with TB by investigating possums in the terminal stage of the disease when the possums have become weak and disoriented and may wander onto pasture during daylight instead of denning (Paterson, 1993; Morris et al., 1994; Paterson and Morris, 1995; Paterson et al., 1995). The probability of cattle becoming infected is dependent on a number of factors that relate to the prevalence of TB in possums at cluster sites, the proximity of cattle grazing areas to possum TB cluster sites, and the length of time over which cattle are exposed to an infective possum (Paterson and Morris, 1995). As a result of these factors there is considerable temporal variation in the annual incidence of TB in cattle and/or deer on farms. We found little evidence of temporal clustering of the annual

128

cumulative incidence of TB in cattle on neighbouring farms across two-year time periods, suggesting that temporal factors affecting the probability of transmission of TB from possums to cattle were not active within the same time period on neighbouring farms (see Chapter 3). Although there was no strong temporal clustering of the annual cumulative incidence of TB in cattle on a year-to-year basis on farms, there was evidence of strong spatial clustering of both the percent of years during which cattle on a farm were infected with TB and the cumulative incidence of TB in cattle over a five-year period from 1986-90, within a VRA located in the south east of the North Island. A possible explanation for this clustering pattern was that habitat patterns influenced the number of persistent possum TB clusters, often referred to as 'hot spots', and the density of possums on farms.
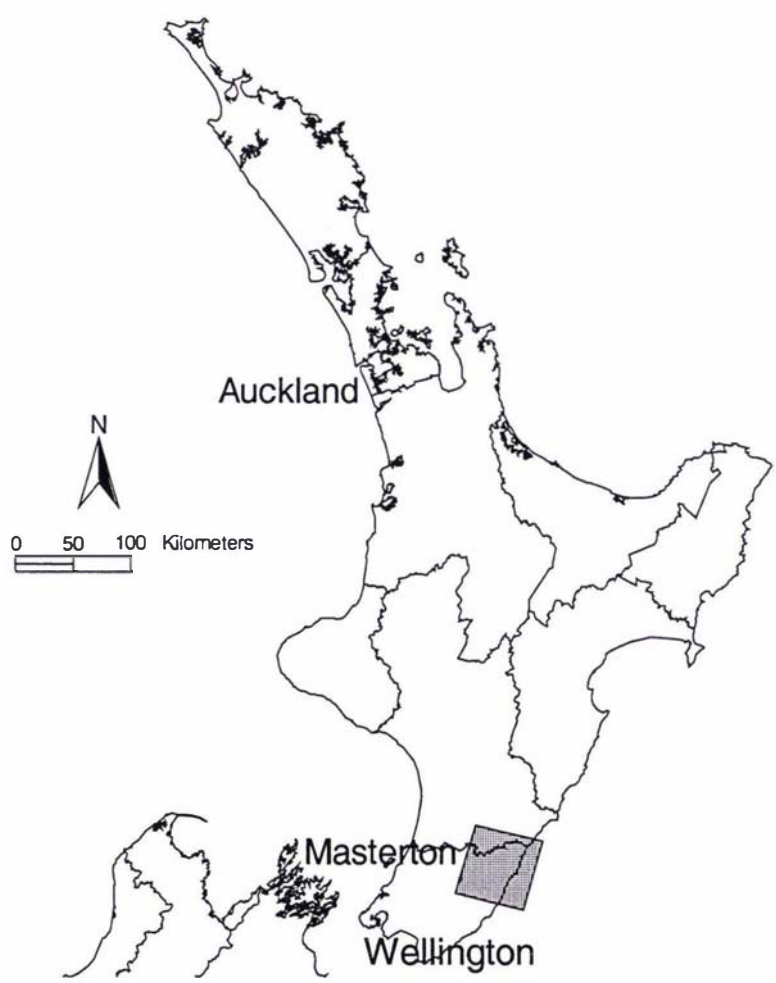
Figure 20. Regional map of the North Island of New Zealand showing the location of the study area.

The association of the distribution of TB in possums with explicit spatial factors such as vegetation patterns makes geographic information systems (GISs) a useful tool to manage and analyse TB data. In recent years GISs have increasingly been applied to the analysis of health data (Hungerford, 1991; Marsh et al., 1991; Carpenter et al., 1994; Norman et al., 1994; Pfeiffer et al., 1994; Rodríguez-Lainz et al., 1996; McKenzie, 1996), particularly vector-borne diseases for which geographical features are important risk factors (Baumann et al., 1994; Glass et al., 1995; Smith et al., 1995; Augustin et al., 1996; Kitron and Kazmierczak, 1997; Kitron et al., 1997; Pfeiffer et al., 1997; Thomson et al., 1998). Remote sensing is increasingly being applied in ecological and vector-borne disease studies (Daniel and Kolar, 1990; Hugh-Jones, 1991; Norval et al., 1991; Rogers, 1991; Perry et al., 1991a; Perry et al., 1991b; Beck et al., 1994; Kitron et al., 1996), providing a cost-effective means of obtaining current vegetation data over a large area (Dymond et al., 1996). McKenzie and Dymond (refer Chapter 4) found that existing digital vegetation data sources in New Zealand did not meet the needs of research into associations between the distribution of possum TB clusters and habitat factors. They produced a more detailed vegetation map from a SPOT3 multispectral satellite image that we used in this study. The recent development of a geographically referenced farm database in New Zealand, Agribase, now makes it possible to collate and analyse geographical data at the farm level (Sanson and Pearson, 1997).

The aim of our study was to identify farm, geographical and possum habitat factors associated with the probability of tuberculous possums being present on farms. In this paper we describe the geographical data used in the study and the geographical analytical techniques used to manipulate these for the analysis. We describe the statistical analytical techniques with particular reference to using cattle TB data clustered at the farm level, and using spatial data in logistic regression modelling. Finally, we discuss the results and their implications for TB management in New Zealand.

Materials and Methods

Study area

The study was conducted in a 60 kilometer square area in the north eastern part of the Wellington region known as the Wairarapa (Figure 20), which corresponded with the coverage of the satellite image used to generate a vegetation map for the study. This is a

pastoral farming area dominated by extensive sheep and beef farming where TB has been endemic in cattle since the 1950s (Shortridge, 1981). TB was first discovered in possums in the Wairarapa in 1968 (Anon, 1993) and since then tuberculous possums have been caught at over 140 different locations throughout the region.

Vegetation in the western part of the study area is predominantly improved pasture species, with patches of manuka/kanuka (*Leptospermum scoparium/Kunzea ericoides*) in gullies, some pine (*Pinus radiata*) plantations and podocarp forest remnants. There are some plantings of exotic species such as willows (*Salix* species) and poplars (*Populus* species) for soil conservation purposes, with willows dominant along river banks. The eastern part of the study area is dominated by a range of hills that runs parallel to the coast and is covered predominantly with pine forest (referred to in the paper as the 'coastal forest'). There are also large areas of beech (*Nothofagus* species) or podocarp/broadleaved forest on the hills, and mixed patches of gorse (*Ulex europeaus*) and manuka occur where the forest has been cleared.

Geographical data

Slope map

We used 20-meter contour lines that had been digitised from 1:50000 topographic maps (Terralink, Wellington, New Zealand) to produce a digital elevation model (DEM) from which we generated a slope map. We initially interpolated height data from the 20-meter contour lines to a resolution of 10 meters and then generalised the resulting DEM to 20-meter resolution. We derived a slope map from the 20-meter DEM using the SLOPE algorithm in Spatial Analyst, ArcView version 3.1 (ESRI, Redlands, California). We then classified the slope data into 5 categories using the RECLASSIFY function. Slope categories were $0-10°$, $11-20°$, $21-30°$, $31-40°$, $>40°$.

Vegetation map

We produced a vegetation map by automatic classification of a SPOT3 multispectral satellite image captured in March 1994 (refer Chapter 4). The resulting map comprised eight vegetation classes and a water class. Vegetation classes included three tree classes (Pine, Podocarp/broadleaf, Beech), four scrub classes (Manuka/kanuka, Manuka/pasture, Manuka/gorse, Shrubland) and a pasture class. Details of the plant species within each vegetation class are presented in Chapter 4. The vegetation data was in 20-meter raster format

which enabled us to use the Spatial Analyst functions for grid maps in ArcView. We did not smooth the data or generalise it to a larger cell size as we wanted to retain the full extent of habitat heterogeneity at a scale that was close to the expected size of possum TB clusters.

River map

We used a river map that had been digitised from NZ topographic maps at a scale of 1:250000 (Terralink, Wellington, New Zealand).

Farm map

We obtained a map of farm boundaries from AgriQuality NZ's national geo-referenced farm database, Agribase (Sanson and Pearson, 1997). This database contains digital farm boundary information plus descriptive information for each farm such as land owner, main enterprise type, farm area and stock numbers. Each farm is identified by a unique farm identification number. The map was not complete for the Wairarapa at the time of the study, thus we were only able to use the farms for which accurate boundary information was available.

Cattle TB data

We obtained cattle and deer TB testing data from the National Livestock Database (NLDB) which is administered by AgriQuality New Zealand (the New Zealand state veterinary service) (Ryan, 1997). This database is designed to manage and record TB testing on farms as required by the national TB control programme. TB data are recorded in the NLDB on a herd basis and records for the Wairarapa date back to 1979. In 1996 the NLDB was amended to enable herds to be linked to the farm on which they were managed by the addition of a unique farm identification number which matched the farm identification number in Agribase. This database has provided a vital link between TB history data and the geographical units from which the data arises, enabling mapping and spatial analysis of TB data to be conducted at the farm level.

Study population

The study population comprised 129 farms on which cattle were raised under a predominantly beef breeding management system and which each had an area greater than 10 hectares. We confined our study to beef breeding management systems to minimise the confounding effect of off-farm sources of TB infection in cattle. Under dry beef management systems almost all cattle are purchased rather than raised on-farm and under dairy management systems many herds are grazed off-farm for a period of the year. As a result the TB incidence in cattle under

both these systems may not be a true reflection of the possum TB situation on the home farm.

Only farms on which the cattle herd had been tested for a minimum of ten years, up to and including the first year of possum control, were included in the study to maximise the chance that the full range of temporal variation in TB incidence was included for each study farm. Fourteen farms in the study group had a deer herd in addition to a cattle herd. Only cattle data were considered on these farms as the deer were generally confined to a small specially fenced area of the farm and we made the assumption that their exclusion did not confound the cattle data.

Outcome variable

Percent TB-positive years was used as the outcome variable. This variable represented the number of years during which at least one TB-positive cattle beast was detected under the national TB surveillance programme as a percent of the number of years that the herd had been under surveillance. TB-positive cattle included all cattle that were recorded in the NLDB as test-positive regardless of whether they had post mortem lesions consistent with TB. Cattle which were test-positive were included in this study as TB-positives regardless of the reason for the testing, which could have been annual herd tests, sale tests, miscellaneous tests, and tests associated with the movement of animals between farms. TB-positive cattle also included cattle that were slaughtered for market purposes and were recorded as having had TB lesions detected during routine inspection at the slaughter plant.

Percent TB-positive years was used in preference to the five-year cumulative incidence for the period 1986-90 as the outcome variable as it was less subject to the annual variability that was associated with cumulative incidence measures. In the case of percent TB-positive years a farm was either negative or positive regardless of the incidence of TB-positive cattle found in any one year or any group of years, and thus it was considered to be a more stable measure of exposure of cattle to vector-derived TB than cumulative incidence. In a preliminary comparison, the spatial distribution of percent TB-positive years appeared to be more strongly associated with vegetation patterns than the five-year cumulative incidence, probably because it was less confounded by cattle management factors and more directly linked to the presence of persistent hot spots.

Independent variables

The unit of interest in this study was a farm plus a 100-meter buffer around its perimeter. We

included the buffer to allow for the influence of possums on a neighbouring property given that the movement of possums is not stopped by farm boundary fences. Spatial variables were calculated using ArcView version 3.1 (Environmental Systems Research Inc, Redlands, California, USA) and its Spatial Analyst extension for grid data. Microsoft Access for Windows version 7.0 (Microsoft Corporation, Redmond, WA, USA) was used to collate the data from the GIS into a format suitable for statistical analysis.

Habitat data

Three different groups of habitat variables were calculated for this study, using the seven non-pasture vegetation categories described above. Pasture was not included in the analyses as it is a constant factor across all farms and it supports extremely low possum populations, thus it has little influence on the number of possums on a farm. A description of the three groups of habitat variables is presented below.

i. Total area of each vegetation class

The total area (in hectares) of each of the seven non-pasture vegetation classes was calculated for each farm and for each 100-meter farm buffer using the TABULATE AREA function. The two values were summed to give a total area of each vegetation class for each farm and its buffer.

ii. Habitat patterns

Habitat pattern variables were calculated using Fragstats, a spatial pattern analysis programme for quantifying habitat structure (McGarigal and Marks, 1995). Fragstats quantifies the areal extent and spatial distribution of habitat patches within a landscape. A PATCH is defined as a contiguous area of cells of the same vegetation type, and is the basic habitat unit used by Fragstats. In this study we used a raster data set with 20-meter cells, thus the minimum size of a patch was 0.04 hectares. A LANDSCAPE is defined as an area of land containing a mosaic of habitat patches, the boundaries of which are dependent on the objective of the study. In this study the landscape was a farm plus a 100-meter buffer around its perimeter. Fragstats calculates variables at three levels of aggregation for each farm: individual patches, individual vegetation classes, and the farm as a whole. Variables are divided into the following groups: area metrics, patch density, patch size and variability metrics, edge metrics, shape metrics, core area metrics, nearest-neighbour metrics, diversity metrics and contagion and interspersion metrics. A list of the variables calculated by Fragstats is presented in Appendix I. As we were only interested in the habitat patterns of non-pasture vegetation classes, we set

pasture as a background variable which excludes it from the calculation of all habitat pattern variables. We use the term COVER in this paper as a general term to refer to the areas of non-pasture vegetation on farms.

Fragstats does not recognise a polygon as defining the boundary of each landscape area of interest, thus we prepared a separate vegetation map for each of the farms included in the study. We used Idrisi for Windows version 2.0 (The IDRISI Project, Clark University, Worcester, MA, U.S.A.) to extract a portion of the original vegetation map that represented a farm plus its buffer, for each farm included in the study. We then ran Fragstats for each map using a DOS batch file and collated the results into an Access database. We selected a subset of the variables output by Fragstats that represented patterns for each vegetation class and summary habitat patterns at the farm level. We did not include individual patch variables as we were interested in patterns aggregated at the farm level.

iii. Area of vegetation-slope categories

Different combinations of vegetation class and slope of the land provide habitat that supports different densities of possums and has different risks of possum TB clusters being present (see Chapter 2). We combined the seven vegetation and five slope categories into a total of 35 vegetation-slope categories. MAP QUERY was used to produce a separate coverage for each vegetation-slope category from the original vegetation and slope maps. The area of each vegetation-slope category for each farm plus its buffer was then calculated using the TABULATE AREA function. The two values were summed to give a total area of each vegetation-slope category on each farm plus its buffer.

Table 13. The TB hot spot risk assigned to each vegetation-slope category.

Vegetation classes	Slope categories (degrees)				
	0-10	11-20	21-30	31-40	>40
Podocarp/broadleaf	Mod	Low	Low	Low	Low
Beech	Low	Mod	Mod	Low	Low
Pine	Mod	Mod	Low	Low	Low
Manuka/pasture	Low	Low	Low	Low	Low
Manuka	Low	Low	Low	Low	Low
Manuka/gorse	Mod	Mod	Mod	High	High
Shrubland	High	Mod	Mod	Mod	Mod

Hot spot areas

Locations of clusters of tuberculous possums are colloquially referred to as 'hot spots'. One of three hot spot risk categories (low, moderate or high) was allocated to each vegetation-slope category based on the probability that the habitat supported by each category contained tuberculous possums. The results of the study described in Chapter 2 and of the longitudinal study conducted by Pfeiffer (1994), Jackson (1995) and Lugton (1997) were used to determine the TB risk classification of each of the vegetation-slope categories, shown in Table 13. The total area of habitat associated with each TB risk category on each farm was identified by summing the area of vegetation-slope categories within each risk category.

Possum number & density

The total number of possums on each farm was estimated by multiplying the area of each vegetation category by the average density of possums per hectare for the category in Table 14. The densities in each vegetation type in the Wairarapa were provided by a person involved in estimating possum densities for control purposes (D. Meenken, Wellington Regional Council, pers. comm.). This was entered into the analysis both as the total number of possums and as density of possums per hectare of total area (farm plus buffer).

Table 14. Possum density per hectare of each vegetation class.

Vegetation category	Possum density (per hectare)
Podocarp/broadleaf	8
Beech	3
Pine	5
Manuka/pasture	2
Manuka/kanuka	3
Manuka/gorse	7
Shrubland	7

River data

The length of river on each farm was estimated by converting the line coverage to a 10-meter grid coverage and then calculating the area of river for each farm and buffer polygon using the TABULATE AREA function. We represented this variable in the analysis both as a total area and as a proportion of the farm plus buffer area.

Distance from coastal forest

A variable for distance from the coastal forest was created as we hypothesised that a higher risk of tuberculous possums may have been associated with proximity to this area because of the high density of possums in the forest. Distance from the coastal forest was used rather than distance from all forests, including large pine plantations, as the coastal forest comprised a heterogeneous mix of pine plantations, areas of gorse and broadleaved scrub, and areas of podocarp/broadleaved forest. This habitat supported large densities of possums, whereas pine plantations outside the coastal forest area were more homogeneous and only supported low densities of possums (Clout, 1977). To calculate this variable polygons were digitised around the areas of coastal forest. The farm shape file was converted to a 50-meter grid file and the DISTANCE function was then used to determine the distance of each 50-meter cell from the coastal forest polygons. The distance of a farm from the forest was calculated by averaging the distance of all cells on each farm.

Farm factors

Total area represented the area of the farm plus the area of the 100-meter buffer in hectares. We calculated a variable to represent the shape of the farm using *Ratio of buffer area to farm area*. We calculated *Livestock density* on the farm using the number of livestock units (LSUs) per hectare, with a multiplication factor of 4 for total cattle and 1 for sheep.

Autoregressive variable (Neighbourhood function)

When using spatial data in statistical modelling it is possible that the assumption of independence of observations is not met as events at one location are likely to be correlated with events at neighbouring locations; that is, there is likely to be spatial autocorrelation in the data (Bailey and Gatrell, 1995). If this pattern is not accounted for in the model there is an increased probability of rejecting the null hypothesis when it is in fact true (Type I error) as the standard error of parameter estimates tends to be smaller. We hypothesised that the probability of tuberculous possums being on a farm may not be independent of the probability on neighbouring farms. To account for this effect a variable that represented the average of percent TB-positive years on contiguous farms was included. This variable was calculated by dividing the total number of TB-positive years by the total number of years tested for all farms that shared a common boundary with a farm, excluding the central farm. Inclusion in the model of a variable that represents the spatial pattern of the outcome variable results in a model referred to as an autoregressive model (Bailey and Gatrell, 1995).

Statistical analysis

The independent variables were divided into four groups on the basis of the nature of the geographical and vegetation information that they represented. The first group comprised general geographical features of a farm with habitat represented as the total area and the proportion of the farm of each vegetation class, and is referred to as the general farm model. The second group comprised variables representing habitat patterns at the farm level and is referred to as the farm-level habitat pattern model. The third group comprised variables representing habitat patterns at the vegetation class level and is referred to as the class-level habitat pattern model. The fourth group comprised variables representing combinations of vegetation and slope data and is referred to as the vegetation-slope model. Analysis of data within these groups enabled us to explore in greater depth the different features of vegetation that were associated with the odds of a farm having at least one TB-positive cattle beast in any one year.

A multi-stage process was used to produce a final statistical model of the data. The first stage involved modelling the variables within each of the four groups. The second stage involved constructing a final model that summarised all four groups of variables. The different groups of variables were modelled with and without the forest distance variable to identify which variables were confounded with distance from the forest. In the case of the vegetation class-level habitat pattern group the first stage involved two steps. In the first step a separate model was constructed for each of the seven vegetation classes and in the second step a summary model for the class-level habitat pattern group of variables was constructed using the significant variables from each individual vegetation class model.

A two-step binomial logistic modelling process was applied at each stage, using the percent years during which TB-positive cattle occurred on a farm as the outcome variable (Curtis et al., 1993; Atwill et al., 1996). The first step involved construction of a statistical model for each group using a stepwise process with ordinary logistic regression, and a p-value of less than 0.05 as the criterion for inclusion and of more than 0.10 for exclusion of a variable. The second step involved random effects modelling to account for the extra-binomial variation between farms (McDermott and Schukken, 1994). The significant variables from the ordinary logistic regression model for each group were initially included in the random effects model. Variables were then removed from the random effects model for each group in a backwards stepping process using a p-value greater than 0.05 as the exclusion criterion.

The final model was constructed from the significant variables in the random effects model for each of the four groups using an initial stepwise ordinary logistic regression process followed by a random effects logistic regression process as described above.

Analysis of residuals

The residuals of the summary random effects logistic regression model were tested for the presence of spatial autocorrelation that had not been accounted for in the model. Firstly, the residuals were mapped to identify if there were any obvious spatial patterns present. Secondly, the residuals were analysed for spatial patterns by computing a variogram using a point coverage of farms, with each farm being represented by the centroid of the largest polygon of the farm.

Table 15: Descriptive statistics for the 129 beef breeding farms included in our study.

Variable	Median	Range
Farm size (ha)	687	76 - 1847
Number of cattle tested	168	9 - 1182
Percent years TB-positive	37	0 - 100
Pine (ha)	2	0 - 170
Beech (ha)	3	0 - 42
Podocarp/broadleaf (ha)	7	0 - 220
Manuka/pasture (ha)	18	0.5 - 73
Manuka/kanuka (ha)	4	0 - 81
Shrubland (ha)	22	0.04 - 154
Manuka/gorse (ha)	4	0 - 76

All regression modelling was conducted using EGRET version 1.02.07 (Statistics and Epidemiology Research Corporation and Cytel Software Corporation, Seattle, Washington, USA). Ordinary logistic regression models were fitted with a quasi-Newton algorithm and random effects models with a modified-Newton algorithm. The goodness-of-fit of each model was evaluated by comparing the scaled deviance for each model with the residual degrees of freedom as described in Curtis et al. (1993). The R^2 of the final model was calculated based on the deviance of the intercept-only model and the deviance of the final model (Mittlboeck and Schemper, 1996). The variogram analysis was conducted using VarioWin version 2.2 (Institute of Mineralogy, BFSH2, University of Lausanne, 1015 Lausanne Switzerland).
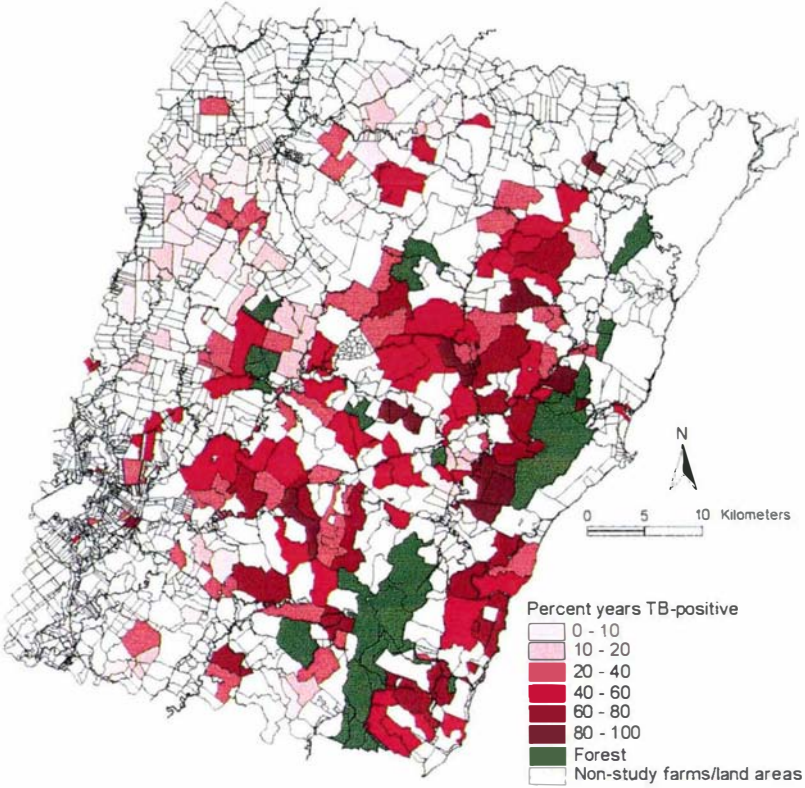
Figure 21. Spatial distribution of percent TB-positive years for the 129 study farms.
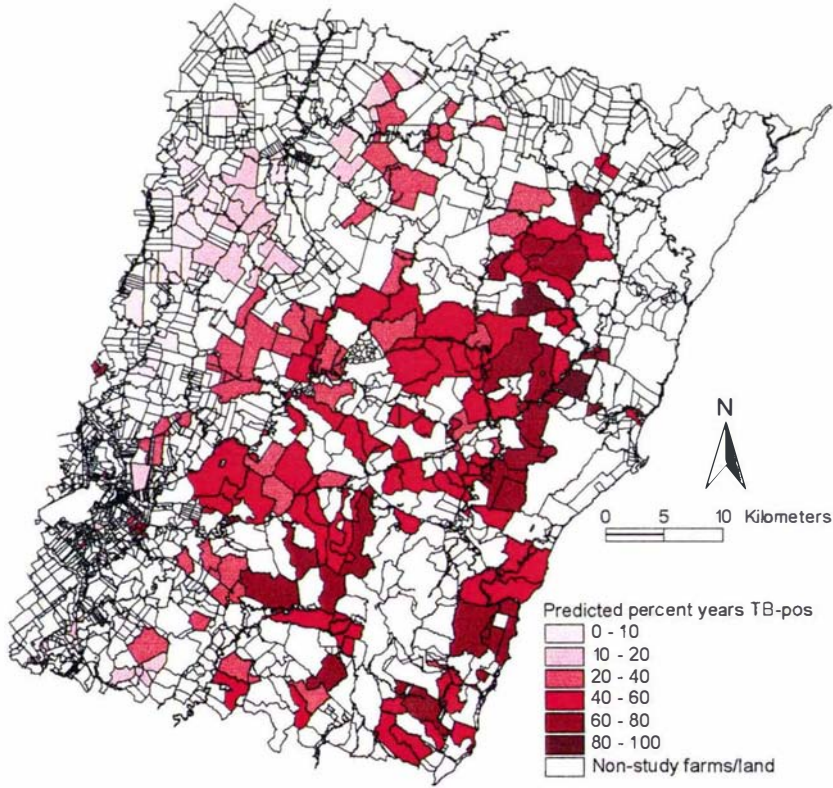
Figure 22. Spatial distribution of predicted percent TB-positive years for the 129 study farms.

Results

We present a summary of the descriptive statistics for study farms in Table 15 and show the distribution of observed and predicted percent TB-positive years for study farms in Figure 21 and Figure 22 respectively.

General farm-level model

We list the variables that we entered into the general farm-level logistic regression model in Table 16, and the variables that were significant in the random effects model with their odds ratio plus 95% confidence limits and associated p-value in Table 17. In the model that did not include Forest Distance, total area of each of the classes Pine, Podocarp/broadleaf and Manuka/pasture and proportion of Beech were positively associated with percent TB-positive years, while the total area of Beech and the proportion of the farm covered by Podocarp/broadleaf were negatively associated with the outcome variable. When Forest Distance was added to the model both the proportion and total area of Beech and the total area of Podocarp/broadleaf became insignificant (Table 17).

Table 16: List of variables entered into the general farm model.

Variable Names		
Total Area (ha)	Area Beech (ha)	Proportion Podocarp/broadleaf
Perimeter (meters)	Area Pine (ha)	Proportion Manuka-Kanuka
Buffer/Farm ratio	Area Podocarp/broadleaf (ha)	Proportion Manuka/pasture
Density livestock units	Area Manuka-Kanuka (ha)	Proportion Shrubland
Forest Distance (2 km)	Area Manuka/pasture (ha)	Proportion Manuka/gorse
Neighbourhood function	Area Shrubland (ha)	Area low hot spot risk (ha)
Area Rivers (ha)	Area Manuka/gorse (ha)	Area moderate hot spot risk (ha)
Number of possums	Proportion Beech	Area high hot spot risk (ha)
Possum density (per ha)	Proportion Pine	

Table 17. Significant variables in the general farm random effects logistic regression model with their odds ratio plus 95% confidence interval, without and with the addition of Forest Distance.

Variable	Model with vegetation variables only		Model with Forest Distance added	
	Odds Ratio	95% C.I.	Odds Ratio	95% C.I.
Area Manuka/pasture (10 ha)	1.37	1.22-1.54	1.22	1.10-1.35
Area Pine (10 ha)	1.32	1.14-1.53	1.15	1.02-1.30
Area Podocarp/broadleaf	1.23	1.00-1.51	1.13	0.95-1.35*
$Prop^n$ Beech	2.23	1.42-3.80	1.04	0.64-1.68*
Area Beech	0.58	0.34-0.99	0.97	0.60-1.58*
$Prop^n$ Podocarp/broadleaf	0.64	0.52-0.78	0.78	0.65-0.93
Forest Distance (2 km)			0.93	0.91-0.95
Deviance (df)	281.73 (121)		242.22 (120)	

* Variables that became insignificant at $p<0.05$ when Forest Distance was included in the model.

Farm-level habitat pattern model

We list the variables that we entered into the farm-level habitat pattern logistic regression model in Table 18, and the variables that were significant in the random effects model with their odds ratio plus 95% confidence limits in Table 19. Total area, patch richness, patch richness density and core area coefficient of variation (CACV) were all positively associated with percent TB-positive years and contagion was negatively associated with the outcome variable when Forest Distance was not included in the model. However, when Forest Distance was included, Total Area was the only variable that remained significant indicating that the effect of the other variables was then represented by the Forest Distance variable.

Table 18: Variables entered into the farm-level habitat pattern model.

	Variable Names	
Total Area (ha)	Landscape Shape Index	Patch Richness
Area Non-Grass (ha)	Area Weighted Mean Shape Index	Patch Richness Density (per 100 ha)
Largest Patch Index (per 1000 ha)	Total Core Area (ha)	Simpson's Evenness Index
Patch Density	Core Area Coefficient of Variation	Interspersion & Juxtaposition
Patch Size Coefficient of Variation	Nearest Neighbour Coefficient of Variation (NNCV)	Contagion
Edge Density	Simpson's Diversity Index	

Patch richness is a measure of species diversity that represents the total number of different vegetation classes (other than pasture) on a farm, regardless of the relative abundance of each patch type or the spatial arrangement of patches. A two-fold increase in the odds of a TB-positive year was associated with every additional vegetation class present on a farm. Patch richness ranged from a minimum of three to a maximum of seven vegetation classes on each farm, and farms towards the east coast tended to have a higher number of vegetation classes than those further inland. Only 10% of farms had 3-5 vegetation classes. These farms had no Pine and very little Manuka or Manuka/gorse. A further 15% of farms had 6 classes. Very few of these farms had Pine and those that did, had only a very small area. Seventy five percent of farms had all seven vegetation classes. Patch richness is generally a function of scale (McGarigal and Marks, 1995). Larger areas are more commonly richer as there is generally greater heterogeneity over larger areas than over comparable smaller areas.

Patch richness density represents the total number of different vegetation classes on a farm divided by the total number of hectares of non-pasture vegetation classes on the farm. This measure standardises patch richness to a per area basis that facilitates comparison among farms. In this study patch richness density was associated with a small increase in the odds of a TB-positive year associated with every extra vegetation class per 100 hectares of non-pasture cover, suggesting a slightly higher risk for farms with smaller areas of cover that were more heterogeneous.

Core area coefficient of variation (CACV) was positively associated with percent TB-positive years. This variable measures the variability in the size of core areas relative to the mean size of core areas of non-pasture vegetation classes on a farm. It is estimated by calculating the standard deviation of core area size as a percentage of the mean size of core areas. Core area is defined as the area within a patch greater than some specified edge distance or buffer width (McGarigal and Marks, 1995). In this study we specified the edge distance as 20 meters. Thus only patches, or parts of patches, with a width and length greater than 40 meters were included as core areas, which meant that rows of trees or very small patches of cover were not included as core areas. CACV tends to be higher for more fragmented or patchy landscapes where the size of patches is more variable. In this study farms with larger areas of heterogeneous cover tended to have higher CACV values as the patches of individual vegetation classes were broken up and interspersed with one another, resulting in more variability amongst patch size than those in areas of more homogeneous cover.

Table 19: Significant variables in the farm-level habitat pattern random effects logistic regression model with their odds ratio plus 95% confidence interval, without and with Forest Distance.

Variable	Model with landscape variables only		Model with Forest Distance added	
	Odds Ratio	95% C.I.	Odds Ratio	95% C.I.
Total Area (100 ha)	1.10	1.04-1.16	1.10	1.05-1.15
Patch Richness	1.91	1.34-2.72	1.02	0.73-1.42*
Patch Richness Density	1.01	1.00-1.02	1.00	0.99-1.01*
Core Area Coefficient of Variation	1.001	1.000-1.002	1.00	0.99-1.00*
Contagion	0.97	0.94-0.99	1.00	0.98-1.02*
Forest Distance (2 kms)			0.92	0.89-0.94
Deviance (df)	298.95 (122)		252.15 (121)	

* Variables that became insignificant at p <0.05 when Forest Distance was included in the model.

Contagion measures the extent to which vegetation classes are aggregated or clumped; higher values of contagion may result from landscapes with a few large, contiguous patches, whereas lower values generally characterise landscapes with many small and dispersed patches (McGarigal and Marks, 1995). Contagion is based on the adjacency of cells, not of patches as is the case with all other landscape variables in Fragstats. The contagion index represents the observed level of contagion as a percentage of the maximum possible, given the total number of patch types. Farms with higher contagion values in this study tended to have more homogeneous areas of vegetation, which commonly comprised Podocarp/broadleaved forest, Shrubland and Manuka/pasture, whereas farms with lower contagion values had more heterogeneous areas of cover in which several vegetation classes were represented within each area.

Vegetation class-level habitat pattern models

We produced a random effects logistic regression model using habitat pattern variables for each of the seven vegetation classes. We list the variables that we entered into each of the vegetation class models in Table 20, and the significant variables in each of the vegetation class-specific random effects logistic regression models in Table 21.

144

Table 20. Variables entered into the logistic regression model for each vegetation class in the class-level habitat pattern models.

Variable Names		
Class area (ha)	Total core area (ha)	Patch Fractal Dimension (PFD)
Number of patches	Total non-core area (ha)	Area Weighted Mean Shape Index (AWMSI)
Patch size coefficient of variation (%)	Percent core area	Nearest neighbour coefficient of variation (%)
Largest patch area (ha)	Number of core areas	Mean Proximity Index
Largest patch index (per 1000 ha)	Core area density (per 100 ha	Interspersion & Juxtaposition Index (%)
Total edge (100 meters)	Mean area of core patches (ha)	

Most of the habitat pattern variables that were positively associated with percent TB-positive years within each vegetation class were associated with areas of heterogeneous vegetation cover that comprise a large number of individual patches or number of core areas, with interspersion of the different vegetation classes. In addition, the larger patches of individual vegetation classes within these areas tended to have an uneven shape as indicated by the variable Area Weighted Mean Shape Index (AWMSI). Variables that were negatively associated with percent TB-positive years represented farms on which the area of largest patch of Podocarp/broadleaf or Manuka/gorse was large relative to the size of the farm, the mean patch size was large relative to the size of the farm, and the density (per 100 hectares of non-pasture vegetation) of core areas of Podocarp/broadleaf, Shrubland and Manuka/pasture was high. Higher density of core areas of these vegetation classes was more likely to occur on smaller farms. Total non-core area of Beech and of Shrubland was negatively associated with percent TB-positive years. Non-core areas were small and/or narrow patches, or parts of patches that were less than or equal to 40 meters wide and 40 meters long. This included single trees or very small patches of non-pasture vegetation and narrow strips of vegetation such as that lining river banks or shelterbelts.

Table 21: Significant variables in each of the class-level habitat pattern random effects logistic regression models, without and with the addition of Forest Distance, with their odds ratio plus 95% confidence limits.

Vegetation Class	Significant Variables	Model with landscape variables only		Model with Forest Distance added	
		O.R.	95% C.I.	O.R.	95% C.I.
Beech	Interspersion index	1.02	1.01-1.02	1.01	0.997-1.01*
	No. core areas	1.27	1.08-1.48	1.18	1.02-1.35
	Total edge (100m)	1.08	1.04-1.11	1.04	1.01-1.07
	Total non-core area (ha)	0.36	0.23-0.59	0.58	0.37-0.89
Pine	Interspersion index	1.02	1.01-1.02	1.01	0.997-1.01*
	Number of patches	1.02	1.01-1.03	1.01	0.999-1.03*
Podocarp-broadleaf	Interspersion index	1.03	1.02-1.05	1.01	0.998-1.02*
	Total non-core area (ha)	1.36	1.20-1.54	1.15	1.03-1.28
	Total edge (100m)	0.98	0.97-0.99	0.99	0.98-0.99
	Largest patch index	0.35	0.21-0.59	0.54	0.35-0.86
	Core area density (per 100 ha)	0.82	0.67-0.99	0.92	0.90-0.94*
Manuka-Kanuka	Number of patches	1.08	1.04-1.12	1.03	0.997-1.06*
	Total non-core area (ha)	2.02	1.25-3.26	1.30	0.87-1.93*
	AWMSI	2.52	1.55-4.09	1.85	1.23-2.78
	Total edge (100m)	0.92	0.87-0.97	0.97	0.93-1.01*
	Largest patch index	0.41	0.17-0.99	0.44	0.21-0.90
Manuka-pasture	Number of patches	1.01	1.005-1.011	1.004	1.002-1.006
	AWMSI	3.03	1.31-7.06	2.33	1.22-4.46
	Core area density (100 ha)	0.76	0.62-0.92	0.83	0.71-0.97
Shrubland	Interspersion index	1.02	1.00-1.03	1.004	0.99-1.02
	Total edge (100m)	1.02	1.01-1.03	1.01	1.00-1.02*
	Total core area	1.16	1.06-1.26	1.08	1.00-1.16
	Total non-core area (ha)	0.86	0.76-0.96	0.93	0.84-1.02*
	Core area density (100 ha)	0.84	0.74-0.95	0.91	0.82-1.01*
Manuka-gorse	Number of patches	1.01	1.003-1.012	1.003	0.999-1.01*
	Interspersion index	1.020	1.005-1.035	1.008	0.99-1.03*
	AWMSI	8.38	2.54-27.59	3.13	1.16-8.49
	PFD	0.03	0.003-0.243	0.25	0.04-1.64*
	Mean patch index	0.92	0.86-0.99	0.98	0.92-1.03*
	Largest patch index	0.18	0.05-0.62	0.29	0.10-0.81

* Variables that became insignificant at p <0.05 when Forest Distance was included in the model.

Table 22: Significant variables in the vegetation class-level random effects logistic regression model that was built from the individual vegetation class models, without and with the addition of Forest Distance.

Variable	Model with landscape variables only		Model with Forest Distance added	
	Odds ratio	95% C.I.	Odds ratio	95% C.I.
Interspersion Manuka/gorse	1.02	1.01-1.03	1.01	0.997-1.02*
Number of patches Manuka/pasture	1.004	1.001-1.007	1.004	1.00-1.01
Number of patches Pine	1.03	1.02-1.04	1.01	0.99-1.02*
Total core area Shrubland	1.07	1.01-1.12	1.04	0.99-1.09*
Core area density Shrubland	0.86	0.78-0.95	0.92	0.83-1.01*
Total edge Podocarp (100 meters)	0.996	0.995-0.998	0.999	0.997-1.001*
Forest Distance (2 kms)			0.93	0.91-0.96
Deviance (df)	279.77 (121)		251.42 (120)	

* Variables that became insignificant at $p<0.05$ when Forest Distance was included in the model.

Addition of the Forest Distance variable to these models showed that many of the variables representing heterogeneity of patches of cover were more likely to occur on farms nearer the forest, likewise the variables representing homogeneity were more likely to occur further inland away from the forest. As a result, their effect became insignificant when Forest Distance was included in the model (Table 21).

Summary vegetation class-level landscape model

When the individual class-level landscape models were combined, many of the variables were no longer significant as there was a high degree of correlation between variables amongst vegetation classes. We list the significant summary class-level landscape variables that were associated with percent TB-positive years in Table 22. Interspersion of Manuka/gorse, number of patches of Pine and number of patches of Manuka/pasture were positively associated with percent TB-positive years, while total edge Podocarp/broadleaf and core area density of Shrubland were negatively associated. Here again, interspersion of Manuka/gorse, number of patches of Pine, total area of Shrubland and core area density of Shrubland were confounded with Forest Distance (Table 22). All of these variables except core area density of Shrubland had higher values on farms closer to the forest. Higher values of core area density of Shrubland were associated with farms further from the forest hence they were negatively associated with percent TB-positive years.

Vegetation-slope model

We list the variables that we entered into the vegetation-slope logistic regression model in Table 23, and the variables that were significant in the random effects model, together with their odds ratio plus 95% confidence limits in Table 24. Total area of Podocarp/broadleaf on flat land (0-10°) was negatively associated with percent TB-positive years. Total area of the farm plus buffer, area of Pine on flat land (0-10°) and area of Shrubland on slopes of 11-20° were positively associated with percent TB-positive years. Pine 0-10° and Shrubland 11-20° were both confounded with Forest Distance and became insignificant once this variable was included in the model (Table 24).

Table 23: List of variables entered into the vegetation-slope model.

Variable Names		
Total Area (ha)	Pod-b'leaf 11-20° (ha)	Manuka/pasture 31-40° (ha)
Beech 0-10° (ha)	Pod-b'leaf 21-30° (ha)	Manuka/pasture 41-60° (ha)
Beech 11-20° (ha)	Pod-b'leaf 31-40° (ha)	Manuka/gorse 0-10° (ha)
Beech 21-30° (ha)	Pod-b'leaf 41-60° (ha)	Manuka/gorse 11-20° (ha)
Beech 31-40° (ha)	Manuka-kanuka 0-10° (ha)	Manuka/gorse 21-30° (ha)
Beech 41-60° (ha)	Manuka-kanuka 11-20° (ha)	Manuka/gorse 31-40° (ha)
Pine 0-10° (ha)	Manuka-kanuka 21-30° (ha)	Manuka/gorse 41-60° (ha)
Pine 11-20° (ha)	Manuka-kanuka 31-40° (ha)	Shrubland 0-10° (ha)
Pine 21-30° (ha)	Manuka-kanuka 41-60° (ha)	Shrubland 11-20° (ha)
Pine 31-40° (ha)	Manuka/pasture 0-10° (ha)	Shrubland 21-30° (ha)
Pine 41-60° (ha)	Manuka/pasture 11-20° (ha)	Shrubland 31-40° (ha)
Pod-b'leaf 0-10° (ha)	Manuka/pasture 21-30° (ha)	Shrubland 41-60° (ha)

148

Table 24: Significant variables in the vegetation-slope random effects logistic regression model with their odds ratio plus 95% confidence interval, without and with Forest Distance added.

Variable	Model with vegetation-slope variables only		Model with Forest Distance added	
	Odds ratio	95% C.I.	Odds ratio	95% C.I.
Total Area (100 ha)	1.14	1.08-1.22	1.10	1.05-1.16
Podocarp 0-10° (ha)	0.84	0.79-0.90	0.94	0.89-0.997
Pine 0-10° (ha)	1.17	1.10-1.24	1.05	0.996-1.11*
Shrubland 11-20° (ha)	1.03	1.01-1.05	1.01	0.996-1.03*
Forest Distance (2 kms)			0.92	0.91-0.94
Deviance (df)	299.56 (123)		247.38 (122)	

* Variables that became insignificant at p <0.05 when Forest Distance was included in the model.

Summary model

We list the significant variables entered into the summary model, that was generated by entering the significant variables from each of the four random effects models into a final logistic model, in Table 25. Total area of the farm plus buffer, total area of Pine and total area of Manuka/pasture were positively associated with percent TB-positive years, while distance from the forest and proportion of the farm covered with Podocarp/broadleaf were negatively associated with percent TB-positive years. This summary model had the highest goodness-of-fit of all five models with a deviance of 241 and 122 degrees of freedom. The model explained 64% of the total variation in the data (R^2 = 0.64). This is made up of 33% which is explained by the biological variables included in the model and 31% which is explained by the random effect variable, which is a variable representing the between farm variation. This is a high level of explanation for data of this complexity.

Table 25: Significant variables in the summary random effects logistic regression model that was built from the significant variables for each of the individual models, with their odds ratio, 95% confidence limits and p-value.

Variable	Odds ratio	95% Confidence Interval	p-value
Forest distance (2 km)	0.93	0.91-0.95	< .001
Total Area (100 ha)	1.05	1.00-1.11	.055
Area Pine (10 ha)	1.13	1.02-1.27	.026
Area Manuka/pasture (10 ha)	1.17	1.04-1.32	.007
Proportion Podocarp	0.88	0.81-0.97	.007
Deviance (df)	241.18 (122)		
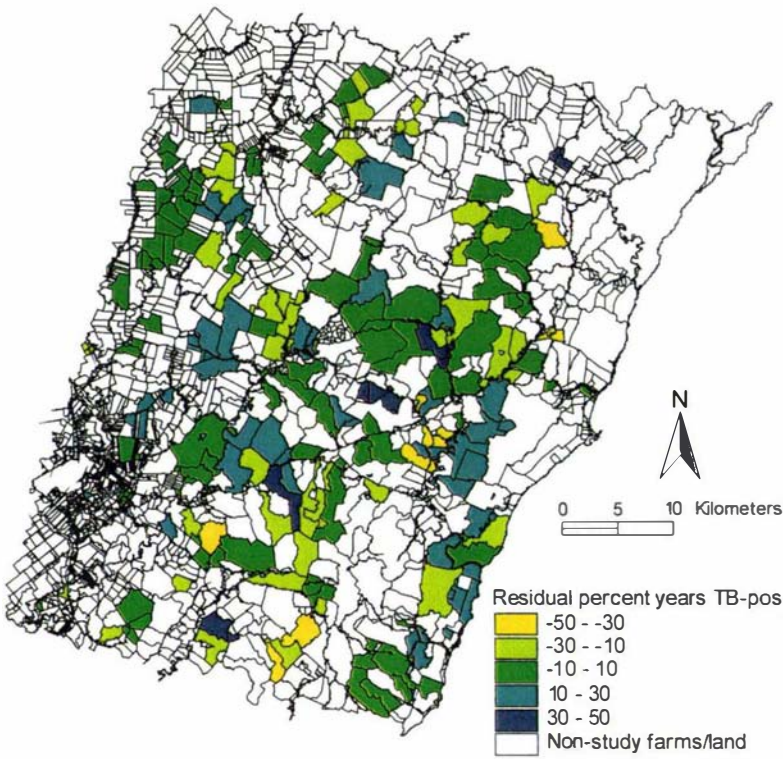
Figure 23. Spatial distribution of residual values from the final random effects logistic regression model.
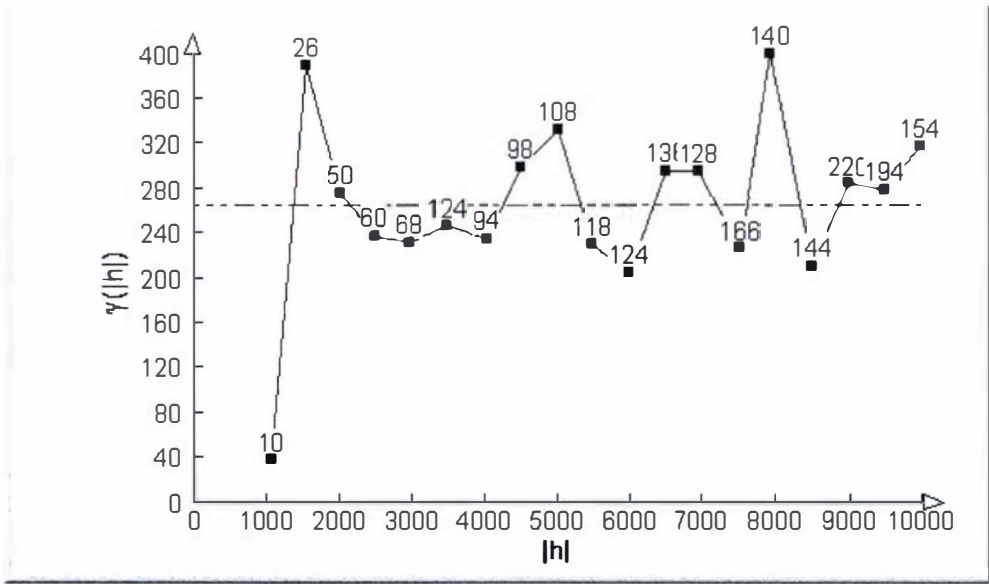
Figure 24. Results of the variogram analysis to test for autocorrelation amongst residuals.

Analysis of residuals

The map of residuals from the summary random effects logistic regression model shows no obvious spatial pattern in residuals on visual examination (Figure 23). The variogram analysis shows that there is a small group of 10 farms whose centroids are within approximately one kilometer of each other that demonstrate some spatial autocorrelation (Figure 24). However, there was no spatial autocorrelation of the residual percent TB-positive years evident for the majority of the study farms, indicating that most spatial autocorrelation in the data had been accounted for by the model.

Discussion

The results of this study show that the most significant factor affecting the odds of a TB-positive year occurring on a farm was distance from the coastal forest (Table 25). A reduction in the odds of a TB-positive year of 7% was associated with every increase of 2 kilometers in the distance from the forest boundary. We believe that this variable was so strongly significant because it represented two separate effects. Firstly, it represented a direct effect of the high possum density within the forest that contributed to the maintenance of high-density possum populations in habitat on neighbouring farms. This high on-farm density may have played an important role in maintaining TB in hot spots on farms. Secondly, it appears to have represented the gradient of heterogeneity of habitat across the study area, which was highest on farms near the forest and lowest on farms furthest from the forest (Figure 25). By running the analyses with and without Forest Distance we were able to demonstrate that it was highly confounded with the spatial pattern of many of the habitat variables such as those representing heterogeneity or homogeneity of cover. The association of proximity to the coastal forest with habitat heterogeneity was most strongly demonstrated in the vegetation class-level habitat pattern models (Table 21, Table 22) in which the variables representing homogeneity and heterogeneity of patches became insignificant once Forest Distance was entered into the models. Because Forest Distance and habitat heterogeneity were strongly negatively correlated it was difficult to entirely distinguish their effects. The density of beef breeding farms was also not uniform with increasing distance from the coastal forest, so the outcome variable (as an indicator of possum TB risk) was not uniform.

Distance from the coastal forest appears to be a crude indicator of multiple habitat factors that are associated with the probability of TB possums occurring on a farm. Inclusion of this

variable results in a model which fits the data used in this study better than models without it. However, its inclusion results in a model that is very site-specific which reduces its predictive value in areas that do not have a forest that is similar to the coastal forest in this study area. While the models that did not include Forest Distance have lower predictive value for the data set used in this study, they provide greater insight into the habitat patterns that are associated with the risk of TB possums being present on a farm and they may have better predictive value for different geographic locations. Of the four models that included different groups of habitat variables, the vegetation class-level habitat pattern model had the lowest deviance when Forest Distance was not included (Table 22) indicating the best fit for the data used in the analysis. This model indicates that farms with cover that comprises Manuka/gorse interspersed with other vegetation classes, a higher number of patches of both Manuka/pasture and Pine and a larger core area of Shrubland have a higher odds of a TB-positive year occurring. Such variables are more likely to occur when the areas of cover on farms comprise more heterogeneous mixes of different vegetation classes compared to those with areas of cover that are more homogeneous and comprise fewer individual patches of Pine and Manuka/pasture with less interspersion of Manuka/gorse. This pattern is illustrated in the vegetation map shown in Figure 25 with a high level of heterogeneity and interspersion of vegetation classes shown in the top right inset and lower levels in the bottom left inset located further from the forest. Figure 26 shows a photograph of an area of very heterogeneous cover with interspersion of many different vegetation classes. Figure 27 shows a photograph of an area of homogeneous Manuka/kanuka cover. Core area represents the area within a patch that is a minimum of 20 metres from the edge of the patch. Thus only patches, or parts of patches, with a width and length greater than 20 metres were included as core areas, which meant that rows of trees or very small patches of cover were not included as core areas. The positive association of total core area of Shrubland indicates that farms with larger patches of Shrubland had a higher risk of TB possums, while the negative association of core area density of Shrubland indicates that smaller farms with larger core areas of Shrubland had a lower risk of TB possums. Likewise farms with a higher total edge of Podocarp/broadleaf patches had a lower risk of TB possums being present.
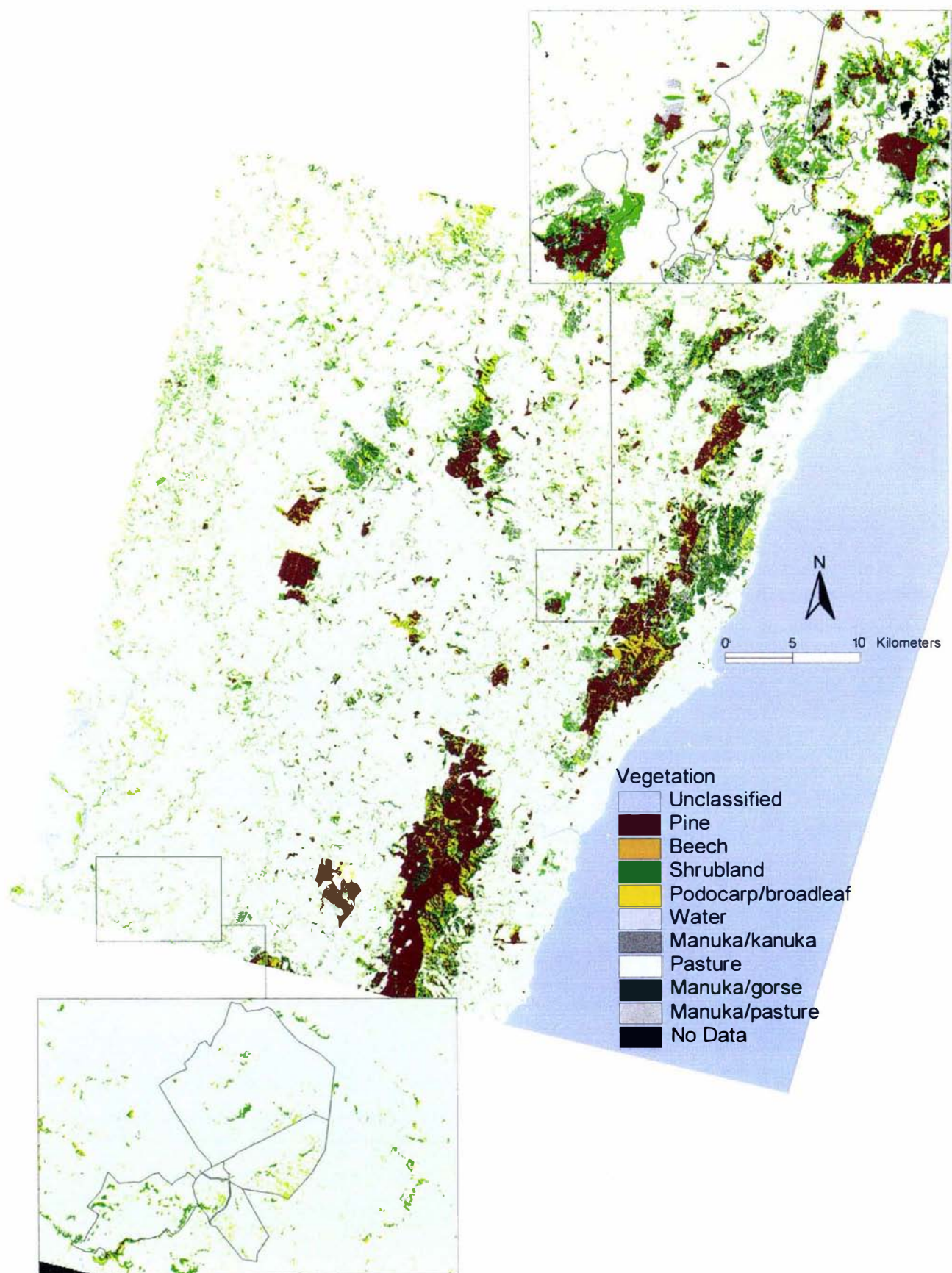
Figure 25. Vegetation map of the study area with insets showing farms with a high degree of heterogeneity (top right) and a high degree of homogeneity (bottom left).

Figure 26. An area of heterogeneous cover with a high degree of patchiness and the interspersion of many vegetation classes.

Figure 27. An area of homogeneous Manuka/kanuka cover.

We evaluated the hypothesis that possum density was an important factor associated with the likelihood of tuberculous possums being present on a farm. Intriguingly, we did not find the density of possums on a farm to be significant in this study. This may have occurred because the Forest Distance variable represented the pattern of possum density more accurately than the density variable that we calculated by estimating the number of possums per hectare of farm plus 100-meter buffer. This density measure is sensitive to farm size, with less stable estimates of possum density occurring on smaller farms. The estimates of possum density on smaller farms were also less reliable indicators of the actual possum density that influences the risk of TB in possums on the smaller farms, as this was more likely to be influenced by possum density over an area extending beyond the area of the farm. It may be more valid to measure possum density within the same area associated with each farm. For future studies we suggest that possum density be measured over a standard area for each farm; for example, within a circle of diameter 3 kilometers (this diameter may be varied to suit local habitat) surrounding the centroid of each farm, to evaluate whether this provides better definition of the influence of possum density within a small locality.

We had expected the area of hot spots, in particular high-risk hot spots, to be significantly associated with percent TB-positive years, as the higher the number of hot spots the higher the chance of at least one tuberculous possum being present on the farm. In this study the area of hot spots on each farm was not significant. This may have occurred because this variable was highly correlated with one or more variables which did enter the model, or because of the inadequate specificity of the hot spot classification system, given the limited differentiation of plant species within the vegetation data. This may have led to over-prediction of hot spots and inclusion of areas of habitat that were not high risk but that were classified into the same vegetation class as other high-risk habitat.

The total area of Pine and of Manuka/pasture on a farm was positively associated with the odds of a TB-positive year occurring. Neither of these vegetation classes supports a high possum density (Clout, 1977; Batcheler and Cowan, 1988), and their significance in this study was most likely due to their being a proxy variable for high-risk patterns of vegetation. For example, farmers tend to plant pine trees in areas where the soil is of poorer quality and the land is covered in scrub so that they can obtain a better return from that land. Thus more pine occurs on farms with more extensive scrub cover, particularly gorse. These variables may have been indicator variables for farms that had more extensive scrub cover, which hence had a higher risk of containing tuberculous possums.

The vegetation classes that were found to be significantly associated with percent TB-positive years in this study, i.e. Pine, Manuka/pasture and Podocarp/broadleaf, were more homogeneous in terms of the range of tree species and/or habitat structure that each class contained, compared with the other vegetation classes (refer to details in Chapter 4). Shrubland and Manuka/gorse were the most diverse classes in terms of the floristic mix found in different localities for the same classified vegetation type. As a result they probably represented different TB risks at different locations despite being classified as the same vegetation class. For example, field checking of the vegetation map showed that many of the areas of cover located in the eastern part of the study area that were classified as Manuka/gorse comprised a mixture of manuka/kanuka and gorse, whereas many of the areas of cover with the same vegetation classification in the western part of the study area comprised manuka/kanuka and broadleaved scrub. The Shrubland class included willows lining riverbanks as well as narrow rows of manuka on pasture, and low-density gorse on pasture. While these were all classified into the same vegetation category the different habitat types can be associated with very different possum densities and may be associated with different risks of supporting possum TB hot spots. We believe that a more detailed vegetation map that had greater differentiation of plant species within these scrub categories would enable more specific vegetation classes to be identified in association with TB risk at the farm level.

This study involved the statistical analysis of spatial data. Farm data in such cases may be spatially correlated due to their relative locations, thus violating the assumption that individual observations are independent (Bailey and Gatrell, 1995; Arlinghaus, 1996). Just as clustering of the units of observation (farm-years) used to measure the outcome variable was handled by including farm as a random effect variable, spatial clustering can be handled by the addition of variables representing the spatial autocorrelation of observations. In this study we incorporated a neighbourhood function variable to represent any correlation between a farm and its immediate neighbours. The lack of significance of the neighbourhood function suggests that any spatial autocorrelation that occurred in the data had been explained by other variables in the model. Alternatively, the proximity measure of contiguity did not represent the scale at which autocorrelation occurred in the data. The neighbourhood function may also have been affected by missing data as neighbours of many farms were missing from the data set because of the incompleteness of the farm boundary map. The cluster of percent TB-positive years described in Chapter 3 showed that spatial autocorrelation occurred at a scale of

several kilometers, and we believe that this effect was picked up by the Forest Distance variable.

Analysis of spatial patterns in the residuals of a logistic regression model is a useful way to determine if spatial autocorrelation in the data has been adequately represented in the model. If there are no significant spatial patterns in residuals we can have more confidence in the accuracy of the parameter estimates in the model. In this study we found no strong evidence of spatial patterns in the residuals based on variogram analysis, suggesting that any spatial autocorrelation in the data was represented in the summary logistic regression model.

In this study we explored different methods of analysing possum habitat and geographical variables at the farm level using a GIS and habitat pattern analyses to identify the variables that would be most useful as predictors of possum TB risk on farms. The variables that were most significant in this study were those associated with the heterogeneity of areas of cover. Farms with more heterogeneous areas of cover had a higher odds of a TB-positive year than those with more homogeneous areas of cover. Having access to a complete map of farm boundaries, from AgriQuality NZ's Agribase is a basic requirement for such analyses and is an essential data set for research and management of wildlife vector-related TB in New Zealand. Further examination of methods for analysing geographical variables, such as that suggested for possum density, is likely to increase the explanatory value of habitat models. There have been significant developments in the remote sensing industry since conducting this study and satellite imagery with higher spectral information and the same spatial resolution is now available, enabling the production of more detailed vegetation maps. As more detailed geographical data becomes available in the future we can use the experience gained in this study to identify factors that will predict the possum TB risk of farms more accurately, providing information to target TB management resources at higher risk farms.

Bibliography

Animal Health Board. (1995). National Tb Strategy: proposed national pest management strategy for bovine tuberculosis. Animal Health Board, Wellington, New Zealand.

Anon. (1986). History of Tb control scheme. In: S. Hennessy (Ed), Surveillance 13 (3): 4-8.

Anon. (1993). Wellington region operational plan for bovine tuberculosis, July 1993 - June 1998. MAF Quality Management, Masterton, New Zealand.

Arlinghaus, S. L., Editor. (1996). Practical handbook of spatial statistics. Boca Raton, Florida,

U.S.A.: CRC Press.

Atwill, E.R., Mohammed, H.O., Lopez, J.W., McCulloch, C.E., and Dubovi, E.J. (1996). Cross-sectional evaluation of environmental, host, and management factors associated with risk of seropositivity to *Erlichia risticii* in horses of New York State. AJVR 57 (3): 278-85.

Augustin, N.H., Mugglestone, M.A., and Buckland, S.T. (1996). An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology 33 : 339-47.

Bailey, T. C., and Gatrell, A. C. (1995). Interactive Spatial Data Analysis. Burnt Mill, Harlow, Essex, CM20 2JE, England: Longman Scientific & Technical.

Batcheler, C.L. and Cowan, P.E. (1988). Review of the status of the possum (*Trichosurus vulpecula*) in New Zealand, for the Agricultural Pest Destruction Council. Ministry of Agriculture and Fisheries, Wellington, New Zealand.

Baumann, M.P.O., Zessin, K.H., and Carlton, T.L. (1994). Application of a geographic information system (GIS) for area-related cross-sectional studies in Sudan - Geographical analytical epidemiology on Trypanosomiasis. In: Proceedings of a WHO consultation on development and application of geographical methods in the epidemiology of zoonoses.

Beck, L.R., Rodriguez, M.H., Dister, S.W., Rodriguez, A.D., Rejmankova, E., Ulloa, A., Meza, R.A., Roberts, D.R., Paris, J.F., Spanner, M.A., Washino, R.K., Hacker, C., and Legters, L.J. (1994). Remote sensing as a landscape epidemiologic tool to identify villages at high risk for malaria transmission. American Journal of Tropical Medicine Hygiene 51 (3): 271-80.

Carpenter, T.E., Snipes, K.P., and Hird, D.W. (1994). A spatial epidemiologic investigation of fowl cholera. In: Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics.

Clout, M. N. (1977). The ecology of the possum (*Trichosurus vulpecula Kerr*) in *Pinus radiata* plantations. Unpublished Ph.D. thesis, University of Auckland.

Curtis, C.R., Mauritsen, R.H., Kass, P.H., Salman, MD., and Erb, H.N. (1993). Ordinary versus random-effects logistic regression for analyzing herd-level calf morbidity and mortality data. Preventive Veterinary Medicine 16 : 207-22.

Daniel, M. and Kolar, J. (1990). Using satellite data to forecast the occurrence of the common tick *Ixodes ricinus* (L.). Journal of Hygiene, Epidemiology, Microbiology and

Immunology 34 (3): 243-52.

Dymond, J.R., Page, M.J., and Brown, L.J. (1996). Large area vegetation mapping in the Gisborne district, New Zealand, from Landsat TM. International Journal of Remote Sensing 17 (2): 263-75.

Glass, G.E., Schwartz, B.S., Morgan, J.M., Johnson, D.T., Noy, P.M., and Israel, E. (1995). Environmental risk factors for Lyme disease identified with geographic information systems. American Journal of Public Health 85 (7): 944-48.

Hickling, G. (1995). Clustering of tuberculosis infection in brushtail possum populations: implications for epidemiological simulation models. In: F. Griffin, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. University of Otago Press, Dunedin, New Zealand. 174-77.

Hickling, G.J. (1989). Assessment of possum control in an area of endemic bovine tuberculosis. Forest Research Institute, Christchurch, New Zealand.

Hickling, G.J. (1991). Ecological aspects of endemic bovine tuberculosis infection in brushtail possum populations in New Zealand. In: Proceedings of the 9th Australian Vertebrate Pest Control Conference. 332-36.

Hugh-Jones, M. (1991). Satellite imaging as a technique for obtaining disease-related data. Review Scientifique Techincal, Office Internationale Epizooties 10 (1): 197-204.

Hungerford, L.L. (1991). Application of spatial statistics to the study of the epidemiology of bovine anaplasmosis. In: Proceedings of the 6th International Symposium on Veterinary Epidemiology and Economics.

Jackson, R. (1995). Transmission of tuberculosis (*Mycobacterium bovis*) by possums. Unpublished PhD thesis, Massey University, New Zealand.

Kitron, U. and Kazmierczak, J.J. (1997). Spatial analysis of the distribution of Lyme disease in Wisconsin. American Journal of Epidemiology 145 (6): 558-66. 47 ref.

Kitron, U., J. Michael, J. Swanson, and L. Haramis. (1997). Spatial analysis of the distribution of LaCrosse encephalitis in Illinois, using a geographic information system and local and global spatial statistics. American Journal of Tropical Medicine & Hygiene 57 (4) 4: 469-75.

Kitron, U., Otieno, L.H., Hungerford, L.L., Odulaja, A., Brigham, W.U., Okello, O.O.,

Joselyn, M., Mohamed-Ahmed, M.M., and Cook, E. (1996). Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. Journal of Animal Ecology 65 (3): 371-80.

Livingstone, P.G. (1997). Update on the New Zealand TB situation. In: Proceedings of a Seminar on Possum and Mustelid Control Research. National Possum Control Agencies, Wellington, New Zealand. 17-30.

Lugton, I. W. (1997). The contribution of wild animals to the epidemiology of tuberculosis in New Zealand. Unpublished PhD thesis, Massey University, New Zealand.

Marsh, W.E., Damrongwatanpokin, T., Larntz, K., and Morrison, R.B. (1991). The use of a geographic information system in an epidemiological study of pseudorabies (Aujeszky's disease) in Minnesota swine herds. Preventive Veterinary Medicine 11 : 249-54.

McDermott, J.J. and Schukken, Y.H. (1994). A review of methods used to adjust for cluster effects in explanatory epidemiological studies of animal populations. Preventive Veterinary Medicine 18 : 155-73.

McGarigal, K. and Marks, B.J. (1995). FRAGSTATS. Spatial pattern analysis program for quantifying landscape structure Ver. 1.0. General Technical Report PNW-GTR-351. Forest Science Department, Oregon State University, Corvallis, Oregon, U.S.A.

McKenzie, J.S. (1996). The application of GIS to animal health issues using TB management and research in New Zealand as an example. In: J. S. McKenzie (Ed), Proceedings of the Epidemiology and State Veterinary Programmes, NZVA/AVA Pan Pacific Conference. Dept of Veterinary Clinical Sciences, Massey University, Palmerston North, New Zealand.

Mittlboeck, M. and Schemper, M. (1996). Explained variation for logistic regression. Statistics in Medicine 15 : 1987-97.

Morris, R.S. and Pfeiffer, D.U. (1995). Directions and issues in bovine tuberculosis epidemiology and control in New Zealand. New Zealand Veterinary Journal 43 (7): 256-65.

Morris, R.S., Pfeiffer, D.U., and Jackson, R. (1994). The epidemiology of *Mycobacterium bovis* infections. Veterinary Microbiology 40 (1): 153-77.

Norman, H.S., Sischo, W.M., Day, R.L., Pitcher, P., and Nesselrodt, A. (1994). Evaluating the influence of neighbouring swine herd status on the spread of pseudorabies virus using

a geographical information system. Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics.

Norval, R.A.I., Perry, B.D., Gebreab, F., and Lessard, P. (1991). East Coast fever: a problem of the future for the horn of Africa? Preventive Veterinary Medicine 10 : 163-72.

Paterson, B. M. (1993). Behavioural patterns of possum and cattle which may facilitate the transmission of tuberculosis. Unpublished Masters thesis, Massey University, New Zealand.

Paterson, B.M. and Morris, R.S. (1995). Interactions between beef cattle and simulated sedated possums on pasture. New Zealand Veterinary Journal 43 (7): 289-93.

Paterson, B.M., Morris, R.S., Weston, J., and Cowan, P. (1995). Foraging and denning pattern of brushtail possum and their possible relationships to contact with cattle and tuberculosis transmission. New Zealand Veterinary Journal 43 (7): 281-88.

Perry, B.D., Kruska, R., Lessard, P., Norval, R.A.I., and Kundert, K. (1991a). Estimating the distribution and abundance of *Rhipicephalus appendiculatus* in Africa. Preventive Veterinary Medicine 11 : 261-68.

Perry, B.D., Norval, R.A.I., Kruska, R.L., Ushewokunze-Obatolu, U., and Booth, T.H. (1991b). Predicting the epidemiology of tick-borne diseases of cattle in Zimbabwe using geographic information systems. In: Proceedings of the 6th International Symposium on Veterinary Epidemiology and Economics.

Pfeiffer, D. U. (1994). The role of a wildlife reservoir in the epidemiology of bovine tuberculosis. Unpublished PhD thesis, Massey University, New Zealand.

Pfeiffer, D.U., Duchateau, L., Kruska, R.L., Ushewokunze-Obatolu, U., and Perry, B.D. (1997). A spatially predictive logistic regression model for occurrence of theileriosis outbreaks in Zimbabwe. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.12.1-12.12.3.

Pfeiffer, D.U., Hickling, G.J., Morris, R.S., Patterson, K.P., Ryan, T.J., and Crews K.B. (1995). The epidemiology of *Mycobacterium bovis* infection in brushtail possums (*Trichosurus vulpecula* Kerr) in the Hauhungaroa Ranges, New Zealand. New Zealand Veterinary Journal 43 (7): 271-80.

Pfeiffer, D.U., Morris R.S., and Sanson, R.L. (1994). Application of GIS in animal disease control - possibilities and limits. In: Proceedings of a WHO consultation on development

and application of geographical methods in the epidemiology of zoonoses.

Rodríguez-Lainz, A., Hird, D.W., Carpenter, T.E., and Read, D.H. (1996). Case-control study of papillomatous digital dermatitis in southern California dairy farms. Preventive Veterinary Medicine 28 (2): 117-31.

Rogers, D.J. (1991). Satellite imagery, tsetse and trypanosomiasis in Africa. Preventive Veterinary Medicine 11: 201-20.

Ryan, T.J. (1997). The New Zealand National Livestock Database. In: Epidemiology Program. Proceedings of the 10th Federation of Asian Veterinary Associations (FAVA) Congress. Cairns, Australia. 167-70.

Sanson, R.L.and Pearson, A.P. (1997). Agribase - A national spatial farm database. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.16.1 - 12.16.3.

Shortridge E.H. (1981). Tuberculosis in cattle and opossums. In: Proceedings of a Veterinary Public Health Seminar. Palmerston North, New Zealand. 83-95.

Smith, T., Charlwood, J.D., Takken, W., Tanner, M., and Spiegelhalter, D.J. (1995). Mapping the densities of malaria vectors within a single village. Acta Tropica 59 : 1-18.

Thomson, M.C., Connor, S.J., Alessandro, U.D., Rowlingson, B., Diggle, P., Cresswell, M., and Greenwood, B.M. (1998). A spatial model of malaria risk in The Gambia: predicting the impact of insecticide treated and untreated bednets on malaria infection. P. Firns (Editor), Proceedings of the 10th Annual Colloquium of the Spatial Information Research Centre. Spatial Information Research Centre, Dunedin, New Zealand. 321-28.

Appendix i

List of the habitat variables output by Fragstats for the three levels of aggregation: patch, vegetation class and landscape (farm).

PATCH INDICES

Patch ID

Area (ha)

Perimeter (m)

Shape index

Core area (ha)

Core area index (%)

Proximity index

CLASS INDICES

Patch type	Class area (ha)
Total area (ha)	Landscape similarity (%)
Largest patch index (%)	Number patches
Patch density (#/100 ha)	Mean patch size (ha)
Patch size standard deviation (ha)	Patch size coefficient of variation (%)
Total edge (m)	Edge density (m/ha)
Contrast-weight edge density (m/ha)	Total edge contrast (%)
Mean edge contrast (%)	Area-weighted mean edge contrast (%)
Landscape shape index	Mean shape index
Area-weighted mean shape	Double log fractal
Mean patch fractal	Area-weighted mean fractal
Core area similarity (%)	Total core area (ha)
Number core areas	Core area density (#/100 ha)
Mean core area 1 (ha)	Core area standard deviation 1 (ha)
Core area CV 1 (%)	Mean core area 2 (ha)
Core area standard deviation 2 (ha)	Core area coefficient of variation 2 (%)
Total core area index (%)	Mean core area index (%)
Mean near neighbour distance (m)	Nearest neighbour standard deviation (m)
Nearest neighbour coefficient of variation (%)	Mean proximity index
Interspersion/juxtaposition (%)	

LANDSCAPE INDICES

Total area (ha)

Largest patch index

Number of patches

Patch density (#/100 ha)

Mean patch size (ha)

Patch size standard deviation (ha)

Patch size coefficient of variation (%)

Total edge (m)

Edge density (m/ha)

Contrast-weight edge density (m/ha)

Total edge contrast index (%)

Mean edge contrast index (%)

Area-weighted mean class edge contrast (%)

Landscape shape index

Mean shape index

Area-weighted mean shape index

Double log fractal dimension

Mean patch fractal dimension

Area-weighted mean fractal dimension

Total core area (ha)

Number of core areas

Core area density (#/100 ha)

Mean core area 1 (ha)

Core area standard deviation 1 (ha)

Core area coefficient of variation 1 (%)

Mean core area 2 (ha)

Core area standard deviation 2 (ha)

Core area coefficient of variation 2 (%)

Total core area index (%)

Mean core area index (%)

Mean nearest neighbour (m)

Nearest neighbour standard deviation (m)

Nearest neighbour coefficient of variation (%)

Mean proximity index

Shannon's diversity index

Simpson's diversity index

Modified Simpson's diversity index

Patch richness

Patch richness density (#/100 ha)

Relative patch richness (%)

Shannon's evenness index

Simpson's evenness index

Modified Simpson's evenness index

Interspersion/juxtaposition index (%)

Contagion (%)

CHAPTER 6

EpiMAN-TB, a spatial decision support system for the management of wildlife-related tuberculosis in cattle and deer in New Zealand

Abstract

EpiMAN-TB is a spatial decision support system that has been developed to assist field veterinarians and pest control managers develop and evaluate TB-related possum control strategies in New Zealand. The basic objective of EpiMAN-TB is to improve the cost-effectiveness of possum control strategies in eradicating TB from possum populations, and preventing the introduction of infection into areas that are free of TB in wildlife populations. The software takes the form of a workbench of tools that include classifying patches of habitat by possum TB risk, classifying farms by possum TB risk, and evaluating possum TB control strategies at the level of individual farms and at regional level. EpiMAN-TB comprises a relational database, map display and spatial analytical tools, simulation models of the spread of TB between possums at both farm and regional levels, and expert systems. It utilises spatial information relating to vegetation cover, topography and farm boundaries, plus non-spatial information relating to TB history and management information for individual farms.

Introduction

Tuberculosis (TB) in cattle and deer is a problem of national concern to the New Zealand pastoral industries due principally to its negative impact on export markets. A national TB control programme has been in place since the 1970s (Anon, 1986; O'Neil and Pharo, 1995). However, efforts to control the disease in farmed animals have been hampered by the presence of TB in wildlife species, the most important being the brushtail possum (*Trichosurus vulpecula*). Infected possum populations are the major source of TB infection for cattle and deer in New Zealand (Jackson, 1995; Morris and Pfeiffer, 1995; Livingstone, 1997). Thus control of the disease in farmed animals involves controlling the disease in infected possum populations.

The control of possums for TB management purposes has become a large business in recent years, principally driven by the Animal Health Board's goal to eradicate TB from New Zealand (Animal Health Board, 1995). A total of $26 million was spent on possum control in the 1996/97 financial year and this is expected to increase to almost $30 million by the year 2000 (O'Neil and Pharo, 1995). The main focus of possum control strategies to date has been the initial reduction of possum abundance by 80-90% over large areas where there is evidence that possums are infected with TB, followed by an on-going programme to maintain the population at less than 40% of its original size. Modelling of Barlow (1991) has shown this to

be the threshold for TB transmission within possum populations. The application of this approach over a large proportion of the areas in which possum populations are infected, known as Vector Risk Areas (VRAs), has produced significant reductions in both the incidence of TB in cattle and the prevalence of herds under quarantine for TB-control purposes (Livingstone, 1997). Despite these achievements problem herds in which persistent reactors occur remain in areas that have been under a wildlife control programme for many years (G. Atkinson and G. Pannett, AgriQuality New Zealand, pers. comm.). Furthermore, historical experience with possum control in New Zealand has shown that continued control is required to maintain possum populations at a low level to prevent the rebuilding of populations and of TB infection through immigration and reproduction of the local population (Anon, 1986; O'Neil and Pharo, 1995).

The task of eradicating TB from possums is far more challenging than that of reducing the risk of transmission of TB from possums to livestock. Morris and Pfeiffer (1995) advised that "no single measure is going to achieve the desired result in controlling wildlife tuberculosis, and the key to progressive success will lie in tighter and tighter integration of a range of measures to reduce the problem to a point where it is no longer a major concern." As disease incidence declines the remaining foci of infection take on greater importance in achieving the goal of eradication. Successful control at this stage requires new strategies that focus on identifying and clearing residual infection of TB in possums. There is a need to target control at different levels, and it is important to complement large-scale control with focussed efforts designed to suit the needs of particular problem areas or farms. Additional information on factors that could be used to differentiate areas with a high risk of TB possums from those with a low risk, both at the habitat patch and the farm level, would enable:

- More cost-effective programmes to be implemented at the larger scale of operational areas, by targeting resources at areas based on possum TB risk during the maintenance phase of control programmes. Such information may speed the rate of reduction of infected possums, and consequently of infected livestock.

- The design of control programmes tailored to the conditions on individual problem farms or problem areas.

- Prevention of the establishment of new Vector Risk Areas (VRAs), defined as areas in which TB is endemic in the possum population.

- Prevention of the introduction of infection into areas from which TB has been eradicated.

There have been many developments in our understanding of the epidemiology of TB in possums in recent years that could be used in the development of possum control strategies for TB control purposes. For example, habitat factors associated with the distribution of possums have been identified at both the possum home range level (refer to Chapter 2) and the farm level (refer to Chapter 5). The explicit association of the distribution of possums and of tuberculous possums with geographic factors makes Geographic Information Systems (GISs) a useful tool for managing and analysing data associated with TB management. With the development of GISs and remote sensing technology in recent years, there is increased availability of digital data sets of geographic variables. For example, McKenzie and Dymond (Chapter 4) produced a vegetation map from a SPOT3 multispectral satellite image for an area in the Wairarapa, demonstrating the application of an automated classification procedure that would be practical for producing detailed vegetation maps for large areas of New Zealand. A national digital database of farm boundaries with associated land ownership and land use information has been developed by AgriQuality New Zealand (Sanson and Pearson, 1997), making possible the mapping of TB data by farm and the aggregation of geographic data to the farm level.

A number of models of the spread of TB within possum populations have been developed. Barlow (1993) and Roberts (1996) have developed deterministic mathematical models that can be used to evaluate possum control strategies at the scale of thousands of hectares. Barlow's model has had a major influence on current possum control strategies implemented to date (Barlow, 1995; Hickling and Efford, 1996). Pfeiffer et al. (1995b) have developed a geographically-based simulation model of the spread of TB amongst possums that can be used to compare the relative effects of different possum control strategies on the prevalence of TB in possum populations at the farm level (Pfeiffer et al., 1995a). Efford (1996) has developed a spatial model of possums by coupling an individual-based demographic model to a raster GIS. As a part of this project we have developed an additional geographically-based model, referred to as TB-Spread, which can be used to evaluate possum control programmes at the regional level (described later in this chapter).

Decision support systems (DSS) provide a very useful way of integrating this epidemiological information, geographic data and simulation models into TB management decisions. In this paper we describe DSSs in more detail and discuss the features of these systems that make their application to TB management in New Zealand relevant. We describe in detail EpiMAN-TB, the spatial DSS that has been developed to assist the development and

evaluation of TB-related possum control strategies in New Zealand. We describe the functions that are available within EpiMAN-TB and discuss how each of these may be applied to develop possum control strategies that make the most cost-effective use of available resources to achieve the long term TB control objectives in New Zealand.

Decision support systems

DSS are defined as "*interactive* computer based systems, which *help* decision makers utilise *data* and *models* to solve *unstructured* problems (Sprague, 1993). He defined the characteristics of a DSS as follows:

- they tend to be aimed at the less well structured, underspecified problems;

- they attempt to combine the use of models or analytic techniques with traditional data access and retrieval functions;

- they specifically focus on features which make them easy to use by non-computer people in an interactive mode; and

- they emphasise flexibility and adaptability to accommodate changes in the environment and the decision making approach of the user."

Cropper and Forte (1997) defined DSS to be:

"Systems which call for, order and promote deliberation and analysis directly relevant to management tasks where complexity and uncertainty make it difficult to arrive at a reasoned response. A distinctive feature of such systems is their use of computer-assisted modelling methods to help make sense of current issues, in exploring options for future policy and action, and in assessing their consequences."

Both of these definitions highlight the focus of DSS on complex tasks that are less well structured and have an element of uncertainty. Decisions relating to the development of TB-related possum control strategies are complex and unstructured as they require judgement and cannot be made according to a strict routine. They need to be targeted to the particular conditions within an area, and what is appropriate for one area may not be appropriate for another. Spatial problems are generically complex and are usually ill-defined or semi-structured (Clarke and Clarke, 1997). The outcome of possum control strategies can be variable. Cropper and Forte (1997) note that when the 'normal' management processes of

deliberation, analysis and decision making can produce outcomes in which managers lack confidence it is normal to try to simplify the decision task by excluding complexity, making unexamined assumptions, or ignoring uncertainties. DSS have most to offer under such circumstances as they enable more of the complexity to be recognized and managed within the decision process.

DSS themselves do not take decisions, but aim to extend the capabilities of the decision maker and not to replace them. They provide for informed decisions by promoting deliberation and analysis of information directly relevant to management tasks (Cropper and Forte, 1997).

Models form an integral part of DSS, as indicated by the dialogue, data and models (D,D,M) paradigm for DSS suggested by Watson & Sprague Jr. (1993). Models are representations of real processes, events and structures expressed as systems of logical or mathematical relationships between variables (Cropper and Forte, 1997). They are abstractions from reality which provide a framework for understanding the nature of a problem and for exploring options for future policy and action without having to test the options in practice through pilot projects or prototype developments. To be of help in decision making, models must capture those characteristics of a decision situation believed to give rise to critical outcomes, or potential points for management intervention: they cannot represent all aspects of the situation, but must capture enough of the complexity to be plausible. Blanning and King (1993) noted that a principal concern with respect to operations research and management science was that it was directed to the construction of models, but there was insufficient attention paid to the implementation of models, and almost no attention paid to the ongoing use of models by practicing managers. This is the case with possum TB models in New Zealand which do not have an interface that makes them easy to use by TB or pest control managers that are unfamiliar with the models. DSS can provide a user friendly interface that enables managers to perform 'what-if' scenarios to explore the outcomes of different decisions.

An important feature of DSS is the development of a user friendly interface which enables users who are not familiar with computers and/or with the analytical processes used in the DSS to have access to the information within the system. Watson and Sprague Jr. (1993) noted that "an appreciation of the importance of the dialog component is gained by recognizing that from the user's perspective, the dialog is the system."

A number of spatial decision support systems have been developed for forest pest

management (Hutchison et al., 1993; Twery et al., 1993; Williams, 1993) and for other areas of environmental resource management (Frysinger et al., 1996; Kessell, 1996). There is increasing interest in New Zealand in DSS for pest control management (McGlinchy, 1996). A number of DSS have been developed to manage animal health problems at both the farm level and the national level (Morris et al., 1993). EpiMAN(FMD) is a DSS that has been developed in New Zealand to manage an outbreak of foot-and-mouth disease (FMD) (Morris et al., 1993; Sanson, 1993). This is a spatial DSS that combines a database management system, a geographic information system, expert system elements, various models of specific aspects of FMD epidemiology, and a statistical analysis capability. EpiMAN-FMD was developed with the aim that the system could be modified and added to, to develop DSSs to help manage outbreaks of other exotic diseases such as EpiMAN-SF (Stärk et al., 1996) and endemic diseases such as EpiMAN-TB (McKenzie and Morris, 1995; McKenzie et al., 1997).

Overview of EpiMAN-TB

EpiMAN-TB is a decision support system designed for the use of field veterinarians and pest control managers, providing them with access to expert knowledge on the epidemiology of TB in possums, TB simulation models, and tools to display and analyse geographic data in an interface that is easy to use. It is a stand alone system designed to run on PCs under Windows 95. It has principally been designed for the use of TB managers within AgriQuality New Zealand (formerly MAF Quality Management) as they are contracted by the Animal Health Board to implement the TB control programme, and they either own or manage the two major databases that contain information used by EpiMAN-TB. At present the database information is extracted from the original two databases and is maintained separately within EpiMAN-TB, which provides flexibility for the DSS to be used by individuals who do not have on-line access to AgriQuality NZ's databases such as private veterinarians or pest control managers. However, the establishment of a live link between EpiMAN-TB and the original databases could be explored in the future as this would provide for more efficient use of computer space and ensure that the information is as current as that in the original databases.

EpiMAN-TB is a spatial DSS that is designed to manage and analyse spatial data. The geographic tools programmed into EpiMAN-TB include spatial data analysis and display functions that are customised to the users' needs as they relate to the development and evaluation of possum control strategies for TB control purposes. We have made a deliberate effort to not be dependent on any particular commercial GIS or database management

software so that the software may be distributed to a number of users without the additional cost of purchasing licences for commercial GIS software. We have aimed to make the design of this system flexible so that it can be refined as further information becomes available in the future. We have also aimed to keep the design as generic as possible so it can be applied to the development of control strategies for other wildlife species that are shown to play a significant role in the maintenance of TB, and can also be adapted to manage other endemic diseases that have a strong spatial component in their epidemiology.

Description of the system

EpiMAN-TB has four main functions that support possum control decisions at two major levels: the individual farm level involving hundreds of hectares and the regional level involving thousands of hectares (Figure 28). These are:

1. Classification of patches of habitat by their risk of supporting tuberculous possums.

2. Evaluation of the effectiveness of control programmes in controlling TB in possum populations at the farm level.

3. Evaluation of the effectiveness of control programmes in controlling TB in possum populations at the regional level.

4. Classification of farms according to the risk of TB possums being present on the farm.

These functions are implemented by combining tools such as a relational database, map display and spatial analysis tools, simulation models of TB in possums at the farm level and at the regional level, and expert systems as illustrated in Figure 28. The basic structure of EpiMAN-TB includes five modules. These include four modules that implement the functions described above plus an administrative module. Development of the five modules has reached different stages. The modules for the administrative functions, the hot spot predictor and the evaluation of farm-level possum control strategies have been completed and tested by end users in the field who have provided feedback on the interface and the functions available. The module for evaluating possum control strategies at the regional level has been implemented and a framework developed for applying different models of the spread of TB within possum populations. The farm risk module requires further research to refine geographic predictors of farm risk before it is developed to an operational stage.
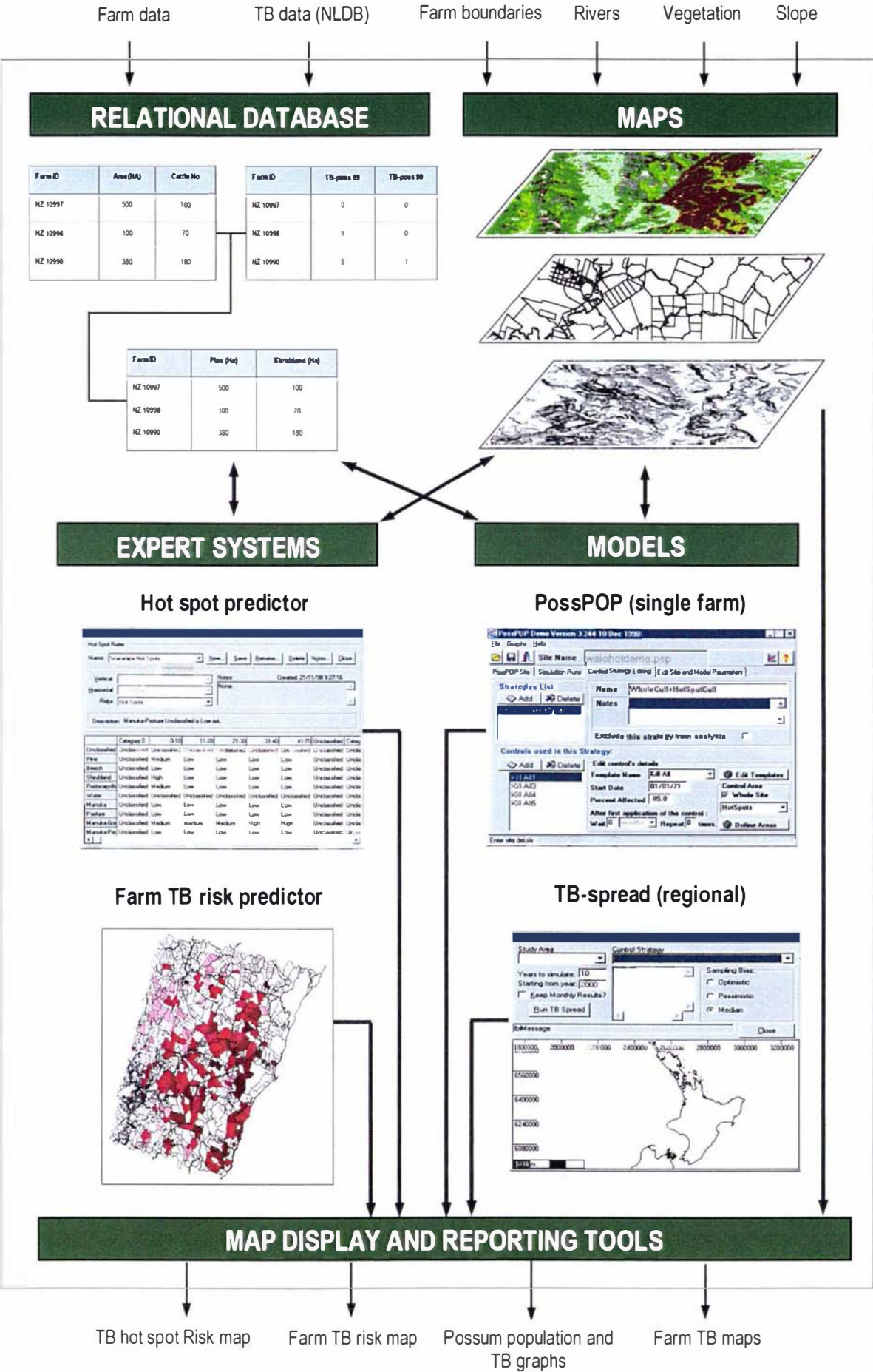
Figure 28. An overview of the structure of EpiMAN-TB

EpiMAN-TB has been developed using geographic data for a 60 kilometre square area in the northeast of the Wellington region, which had been obtained for research into habitat and topographic factors associated with the distribution of TB possums at the farm level (refer to Chapter 5). More extensive data sets can be incorporated as the system is refined and adopted for operational purposes.

Spatial data

EpiMAN-TB has been designed to use spatial data in both vector and raster data format. The system does not include a wide range of sophisticated spatial manipulation tools and users require access to map data that has already been processed with a GIS into the form required by EpiMAN-TB. Data sets currently included in EpiMAN-TB include slope, vegetation, rivers, roads, coastline, and farm boundaries. These maps are described in more detail below.

Slope map

Slope data is represented in five categories: 0-10°, 11-20°, 21-30°, 31-40°, >40°, in a raster format with a spatial resolution of 20 metres. The digital elevation model (DEM) that was used to generate the slope map for EpiMAN-TB was produced by interpolating height data from 20-metre contour lines to a resolution of 10 metres and then generalising the resulting DEM to 20-metre resolution. The 20-metre contour lines had been digitised from 1:50000 topographic maps (Terralink, Wellington, New Zealand). The slope map was generated using the slope algorithm in Spatial Analyst, ArcView version 3.1 (ESRI, Redlands, California).

Vegetation map

Vegetation data is represented in eight categories that include three tree classes (Pine, Podocarp/broadleaf, Beech), four scrub classes (Manuka/kanuka, Manuka/pasture, Manuka/gorse, Shrubland) and a pasture class. Details of the species composition of each vegetation class are presented in Chapter 4 of this thesis. The vegetation data was produced by automatic classification of a SPOT3 multispectral satellite image that had been captured in March 1994 (McKenzie and Dymond, Chapter 4) and is in raster format with a resolution of 20 metres.

Rivers map

River data was digitised from NZ topographic maps at a scale of 1:250,000 (Terralink, Wellington, New Zealand), and is in vector format.

Farm map

Farm boundary data was obtained from AgriQuality NZ's national geo-referenced farm database, Agribase (Sanson and Pearson, 1997). This database contains digital farm boundary information in vector format plus descriptive information for each farm such as land owner, main enterprise type, farm area and stock numbers. Each farm is identified by a unique farm identification number.

Non-spatial data

Non-spatial database information required to run EpiMAN-TB relates to farm ownership, animal numbers and TB status of cattle and deer on the farms. This information is currently available in databases which are either owned or managed by AgriQuality New Zealand. Farm information is obtained from Agribase which is a national database of farms in which each farm is uniquely identified. Agribase contains basic property ownership and land use information plus locational information that facilitates the production of farm maps. TB status information is obtained from the National Livestock Database (NLDB). This database contains a history of TB testing results for most farms in New Zealand. Farms are identified by the Agribase farm identification number so that information in the two databases can be linked.

Functions

Users are able to select from a number of different tasks available within the software by means of a simple interface. The initial screen shows the five main modules included within EpiMAN-TB (Figure 29). These include the four main functions of the DSS plus an administration module. Each module is described in more detail below.
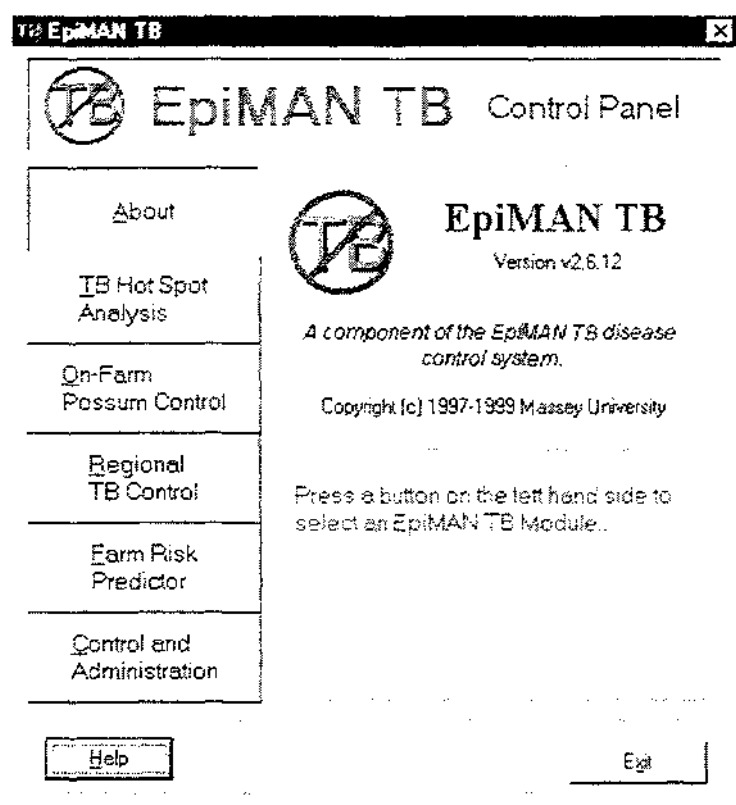
Figure 29. Opening screen of EpiMAN-TB.

TB hot spot analysis module

"Hot spot" is the colloquial term used to describe a location where a cluster of tuberculous possums is believed to exist. Hot spots are critical to the eradication of TB from possum populations as they are the most likely location where tuberculous possums persist following population control programmes. The ability to predict the risk of a patch of habitat being a TB hot spot enables possum control efforts to be refined so that high-risk areas of habitat can be targeted more intensively than low-risk areas. This prediction system is useful for the development of possum control programmes at both the regional scale and the individual farm scale. At the individual farm scale, having the ability to predict the location of areas of habitat where TB is likely to be persistent in possums helps develop a TB management plan for the farm. These high-risk areas can be targeted for more intensive possum control efforts, and/or can be avoided in a cattle or deer grazing programme. Having the information to develop such a targeted programme is more likely to motivate a farmer to implement the programme and also is likely to increase the probability of eradicating TB from the possum population on the farm. Pest managers responsible for possum control programmes at a large scale could use the

information to help refine their programmes, for example by targeting hot spots more frequently or with more intensive possum control methods compared to those applied to low-risk patches of habitat.

The functions available within the hot spot module are shown in Figure 30 and are described in more detail below.
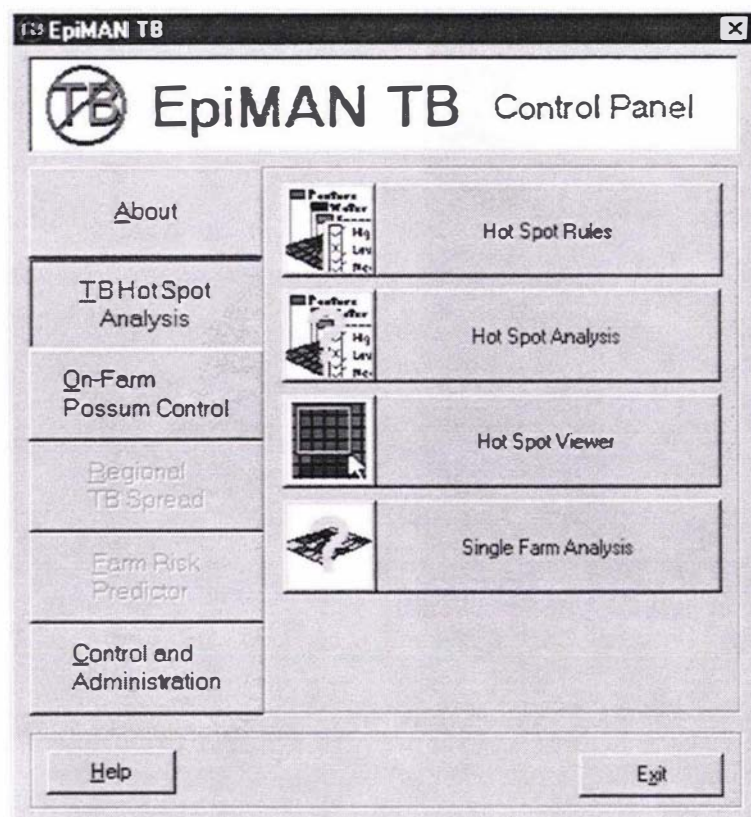
Figure 30. Functions available within the hot spot predictor module.

Hot spot rules

The core of the hot spot module is a set of rules (a simple expert system) for allocating one of three hot spot risk categories (high, medium or low) to each cell of a grid using a combination of vegetation and slope information. The high-risk category represents locations where clusters of TB possums are likely to persist over many years as they provide conditions that favour the transmission of TB between possums. The medium-risk category represents locations where tuberculous possums may be located but on a more sporadic basis than at high risk locations as they provide conditions that are less favourable for the transmission of TB between possums. The low-risk category represents locations where TB possums are unlikely to be located at any time. Vegetation and slope factors that combine to produce high

risk hot spots may differ between different geographical areas, thus users require the flexibility to define and store different rule sets for different areas. Users are able to create a set of hot spot rules customised to their area of interest by means of a form that is shown in Figure 31. Vegetation and slope categories are presented in a simple matrix, and a hot spot risk is allocated by clicking on each cell in the matrix. We developed the set of rules for the Wairarapa shown in Figure 31 based on combined information from the hot spot study (see Chapter 2) and the longitudinal study conducted in the Wairarapa (Pfeiffer 1994; Jackson 1995).

	Category 0	0-10	11-20	21-30	31-40	41-70	Unclassified	Categ
Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Uncla
Pine	Unclassified	Medium	Low	Low	Low	Low	Unclassified	Uncla
Beech	Unclassified	Low	Low	Low	Low	Low	Unclassified	Uncla
Shrubland	Unclassified	High	Low	Low	Low	Low	Unclassified	Uncla
Podocarp-Bi	Unclassified	Medium	Low	Low	Low	Low	Unclassified	Uncla
Water	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Uncla
Manuka	Unclassified	Low	Low	Low	Low	Low	Unclassified	Uncla
Pasture	Unclassified	Low	Low	Low	Low	Low	Unclassified	Uncla
Manuka-Goi	Unclassified	Medium	Medium	Medium	High	High	Unclassified	Uncla
Manuka-Pa:	Unclassified	Low	Low	Low	Low	Low	Unclassified	Uncla

Figure 31. Screen for entering hot spot rules associated with each combination of vegetation and slope category.
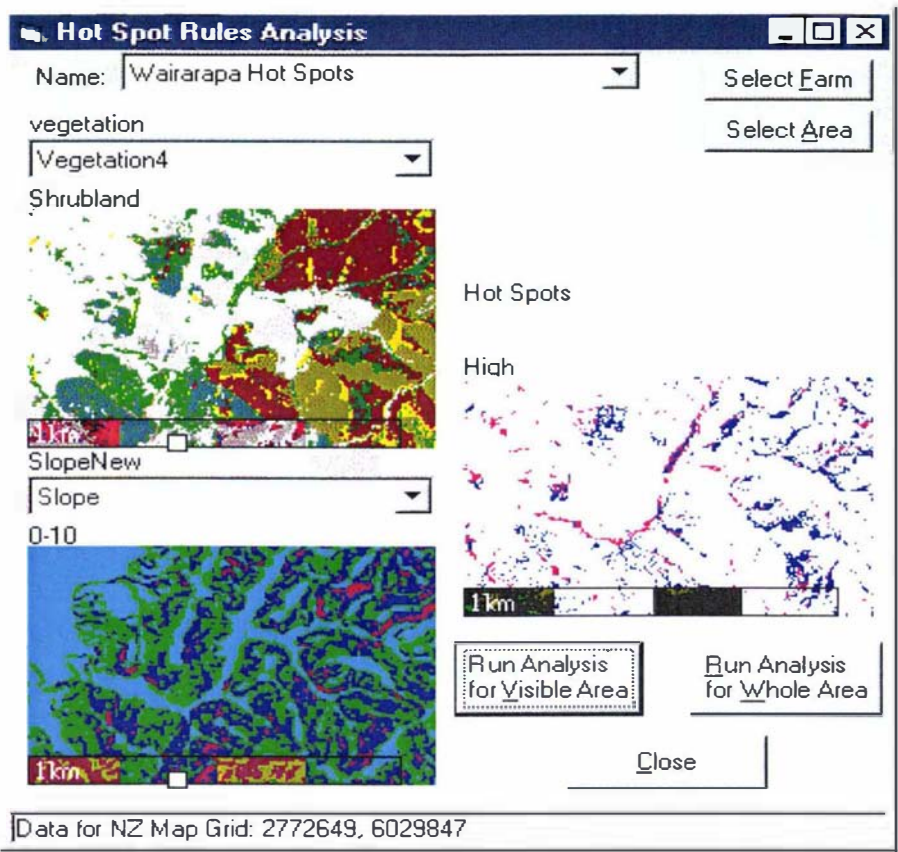
Figure 32. Hot spot analysis screen which enables the user to view vegetation, slope and hot spot data simultaneously.

Hot spot analysis

The hot spot analysis screen enables users to view in detail the input vegetation and slope maps and the resulting hot spot risk map for a small area simultaneously on the same screen before running the analysis for a larger area (Figure 32).

Hot spot viewer

This is the main map viewing tool available in EpiMAN-TB (Figure 33). Any of the geographic data that has been imported into the DSS can be viewed using this tool. All layers that have been loaded for viewing can be switched on or off, enabling the user to choose which layer to view. Vector data can be overlaid on raster data, which enables the user to identify data associated with a specific farm or a specific pre-defined geographic area. An extensive menu system is available to access boundary data for a specific farm in different ways. For example, by means of entering the farm identification number if it is known; alternatively entering the name of the land owner, or entering the name of the road on which the farm is located, which will bring up a list of all farms located on that road.
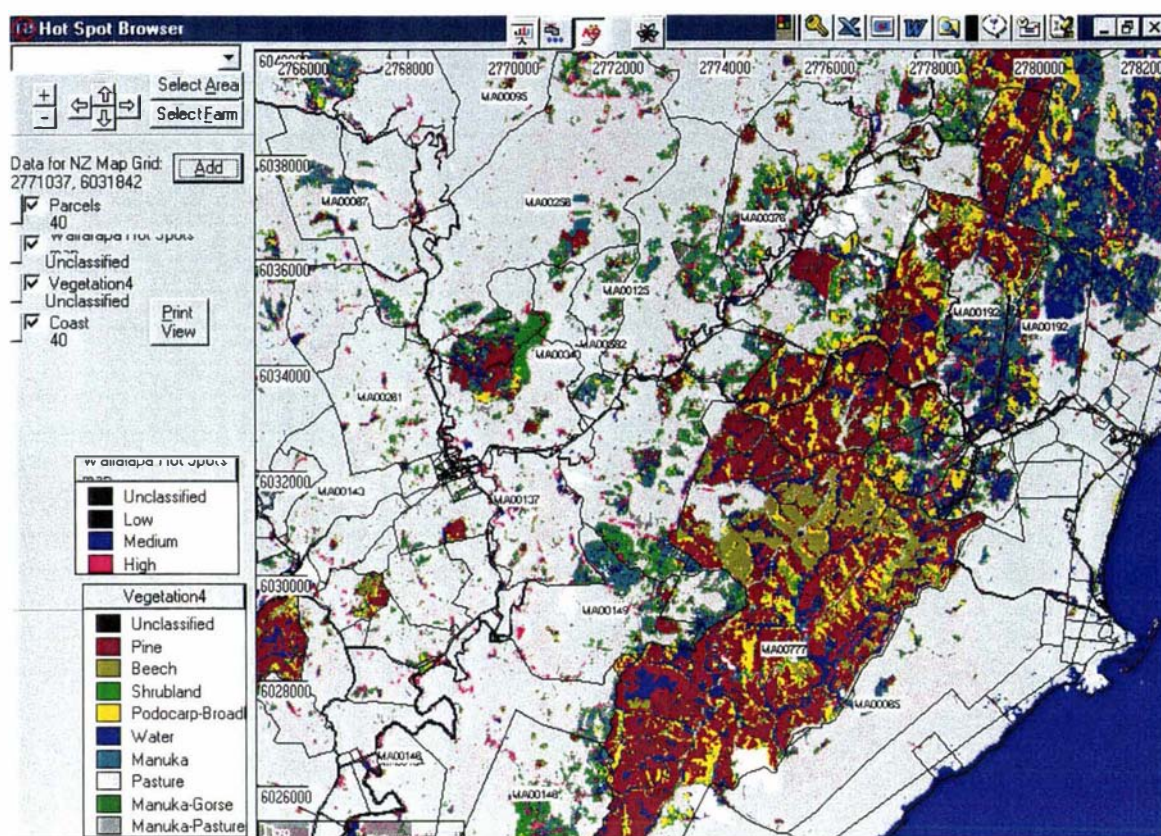
Figure 33. Map viewer function in which all spatial data in EpiMAN-TB can be viewed.

A printing function provides users with the option to print views of data as paper maps.

Single farm analysis

The single farm analysis facility includes tools to edit raster data and to store the edited data in a separate file so that the original database is not modified. This function has not yet been developed. However, it has been included in the concept as it was a function that TB managers who trialled the software wanted included. This enables users to modify the hot spot or the vegetation classification of a patch of habitat having obtained further on-the-ground information from the farm concerned.
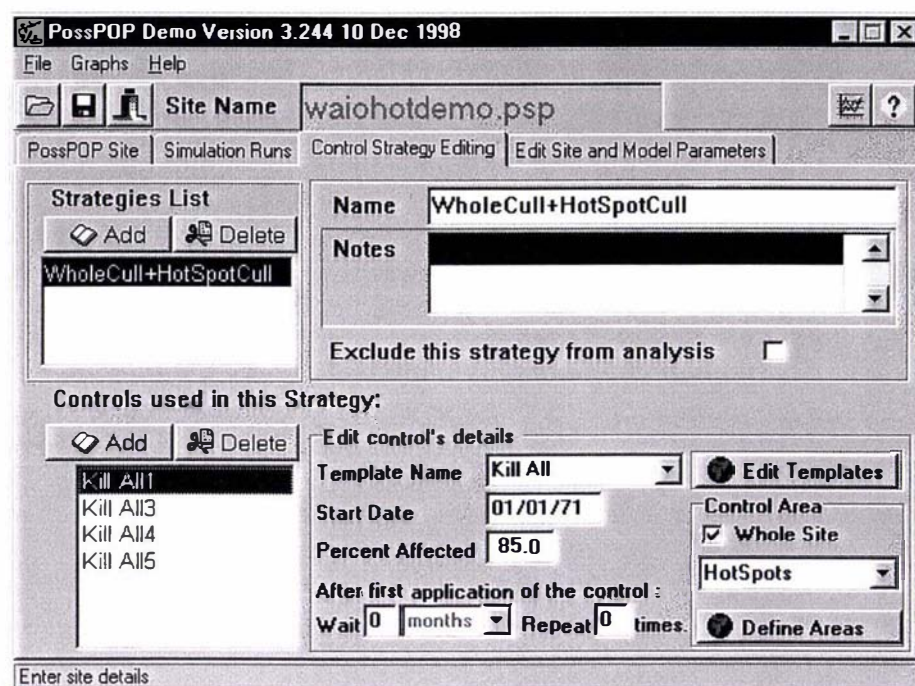
Figure 34. Screen for entering the possum TB control strategies to be modelled in PossPOP.

On-farm possum control module

This module is used to evaluate the effect of different possum control strategies on the prevalence of TB in possums at the scale of an individual farm or a small user-defined area. The module incorporates a simulation model of TB in possums, known as PossPOP. PossPOP is a geographic model representing the ecology and infection dynamics of wild possum populations, which includes natural stochastic variation, spatial patterns such as spatial heterogeneity and autocorrelation, and temporal patterns representing seasonal and cyclical effects (Pfeiffer et al, 1994). PossPOP uses a real vegetation map to "populate" the model with both possums and possum den sites for the area of interest in the simulation. The model can also use the hot spot risk map produced by the hot spot predictor model to adjust the probability of TB transmission between possums in accordance with the vegetation cover. When a hot spot map is included in the model the probability of transmission between possums is modified such that it is higher within hot spot areas than without.

Parameters associated with possum control programmes that can be manipulated by the user include: percent reduction in population, frequency and duration of population reduction, and the location over which population reduction is applied (Figure 34). Users may digitise polygons on-screen to delineate sub-areas within the larger area being modelled, in which a

different control programme may be defined. This enables habitat risk factors to be taken into consideration in the design of control programmes. For example, a programme with the same level of possum reduction over the entire farm can be compared with a programme that has a higher and more frequent population reduction programme in high and medium-risk patches of habitat compared to low-risk areas. Figure 35 shows an example in which individual control areas have been digitised on-screen by the user to model a programme that involves a higher frequency of control in hot spot areas compared to the surrounding area. The graphical output provided by PossPOP includes possum population parameters, TB infection parameters, and location of 'infectious' den sites.
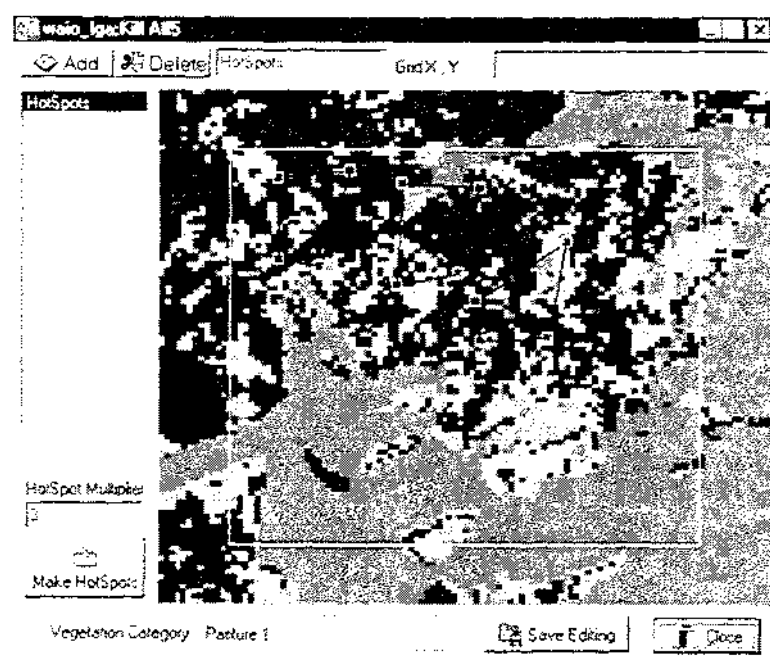
Figure 35. Screen shot from PossPOP showing sub-areas defined by the user which have a different control programme compared to the surrounding area.
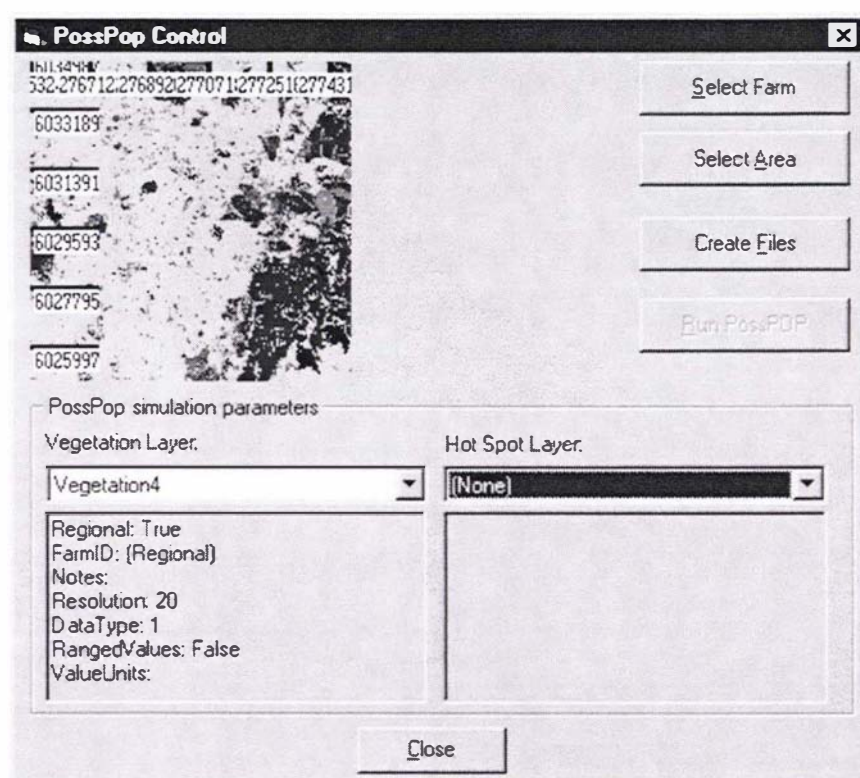
Figure 36. Screen for defining the vegetation and hot spot maps to run PossPOP.

EpiMAN-TB provides an interface that allows the user to define the vegetation map and the hot spot risk map for the area of interest and then to run the model (Figure 36). As for the hot spot prediction model, the geographic boundaries of the vegetation map can be defined either by entering a farm identification number or by an interactive process. If the user wishes to include the habitat risk map in the model, its geographic boundaries can be defined in the same way.

Regional TB spread module

Most possum control programmes are implemented at the scale of several thousand hectares (Anon, 1993), to counter the problem of immigrating possums maintaining TB infection in the controlled area. The size of the area chosen for control in any one operation is influenced by a number of factors that include the extent of contiguous farms with a high rate of TB in cattle, the location of geographical features that may prevent or inhibit the movement of possums (such as rivers and mountain ranges), and resource constraints. Mathematical models have been developed and used in the development of possum control strategies over this large scale in New Zealand (Barlow, 1991; Roberts, 1996). However, these models are run using a hypothetical possum population that has no explicit association with space. We have

developed a model, known as TB-Spread, to evaluate the effects of control programmes at the regional level by incorporating spatially explicit features such as the boundaries of operational areas, farm boundaries and cattle TB data, plus actual vegetation and topographic data. This enables the mosaic of different control programmes that are implemented within different areas of a region, referred to as 'operational areas', to be modelled simultaneously over time.

TB-Spread is a geographic simulation model of the spread of TB at a regional scale. It is based on 25-hectare square land units which are generated by overlaying a vector coverage in the form of a 25-hectare grid on a vegetation map. Each square polygon representing a land unit is given a unique identification number, and the area of each vegetation class within each land unit is tabulated and used to calculate various parameters associated with land units.

This is a very flexible model design that embeds sub-models within a framework that uses a database as the integration and communication mechanism between sub-models, such as that described by Watson and Sprague Jr. (1993). The model centers around a database which contains data on each land unit and has a set of sub-routines that extract data from the database, process it through a user- defined model, and then feed results back to the database. A schematic representation of the model is presented in Figure 36, showing the different components of data included in the database that is central to the model. The raw geographic and topographic data plus the intersecting farm polygons associated with each land unit are represented in the left hand column, and the initial population size and TB prevalence for possums in each land unit is represented in the central column. The dynamic component of the model is represented in the right hand column, in which different models may be used to represent the change in possum population and TB prevalence over time, given a specified control programme. This framework assumes that although simulation models might operate in very different ways they require and produce the same basic set of information; that is, total number of possums and prevalence of clinical and subclinical TB within the possums. The differences between models generally relate to the predicted levels of each of these variables, rather than to the variables that are used.

186
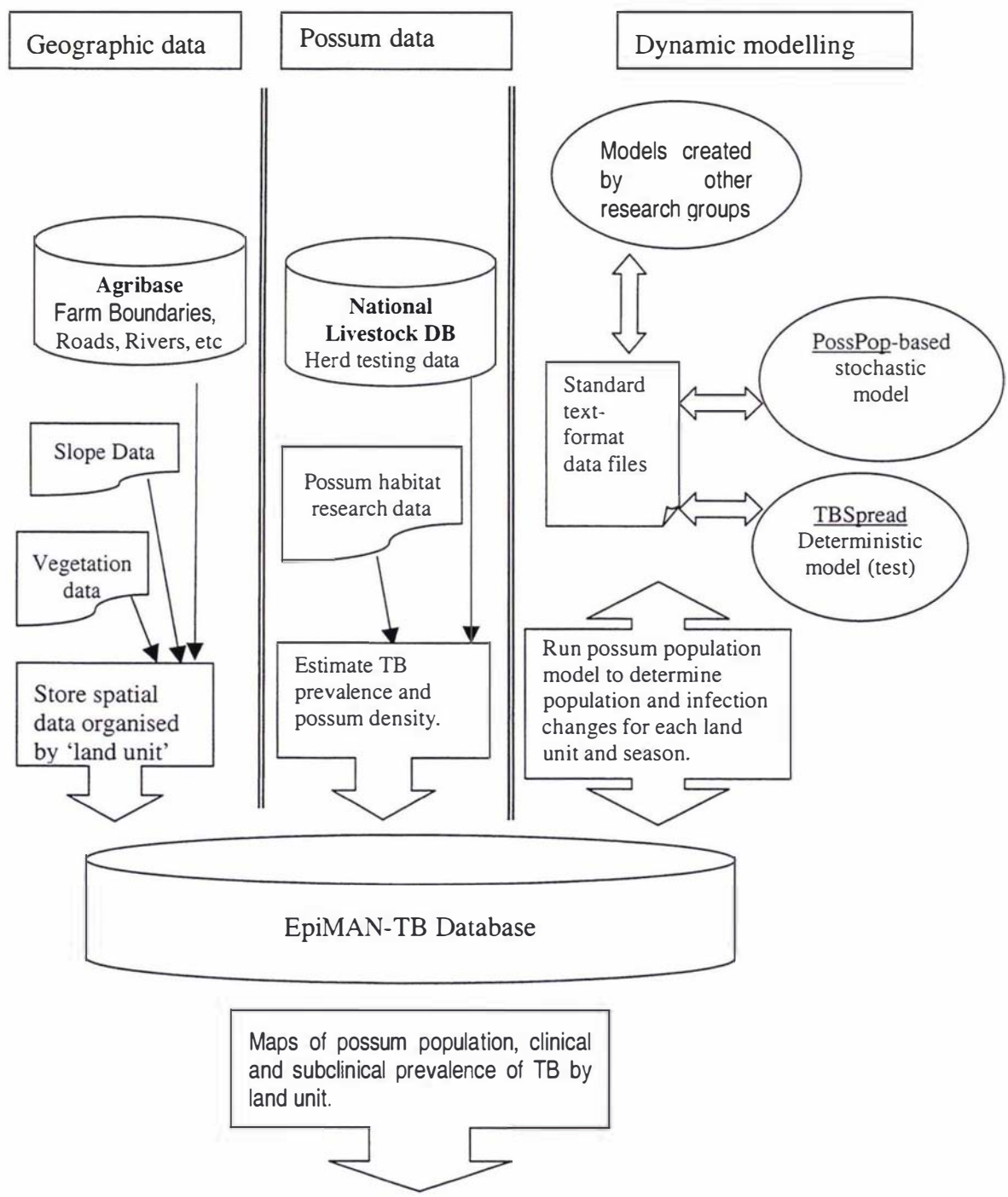
Figure 37. Schematic diagram of TB-Spread.

Notes.
Although the diagram shows TBSpread and PossPop-based models operating via text files these programmes actually interact directly to a set of tables in the database. The text-file interface is available for existing programs which cannot easily be modified to interact with the database. The same information is involved in both cases.

The components of TB-Spread are shown in Figure 38, and each is described in more detail below.
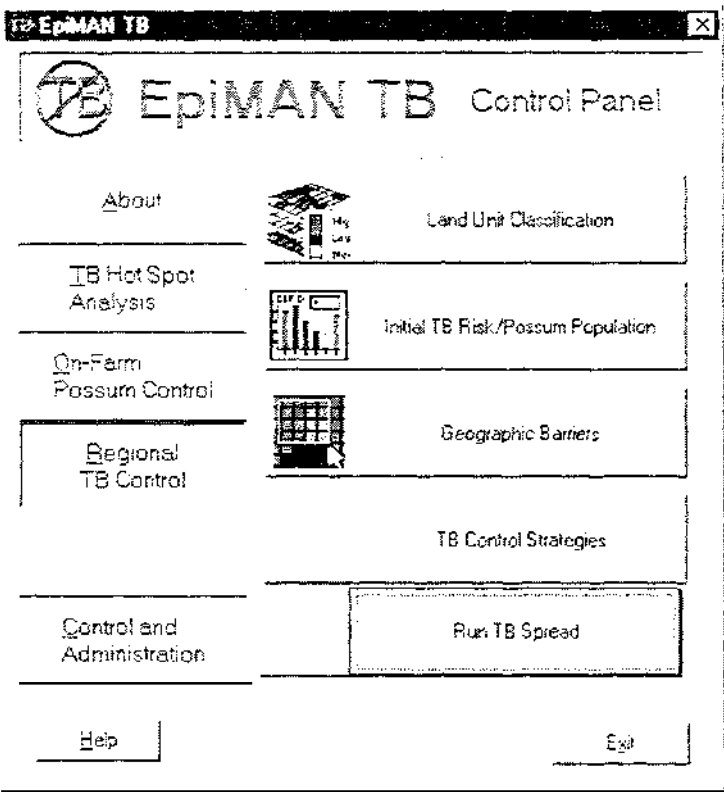
Figure 38. Components of the TB-Spread module.

Vegetation classification

Each land unit is assigned a vegetation category based on rules that relate to the vegetation pattern within each land unit. Currently a very simple set of categories has been developed to illustrate this function (Table 26). However, these can be more complex if desired.

Table 26. Description of the vegetation patterns within each land unit vegetation category.

Land unit vegetation category	Details of vegetation patterns within each land unit category
Pasture	> 80% pasture
Pine	> 60% pine
Mixed	Any combination of vegetation classes other than the two described above.

Initial possum population

The area of each vegetation class is used to populate each land unit with possums by multiplying the area of each vegetation category by the number of possums per hectare for each category, giving a total number of possums per land unit. Possum density figures used for each vegetation category are listed in Table 27.

Table 27. Possum density per hectare of each vegetation class.

Vegetation category	Possum density (per hectare)
Podocarp/broadleaf	8
Beech	3
Pine	5
Manuka/Pasture	2
Manuka/kanuka	3
Manuka/Gorse	7
Shrubland	7

Initial TB prevalence in possums

The initial TB prevalence in possums in each land unit is derived from the incidence of TB in cattle. This is achieved by overlaying the farm boundary map on the land unit map. A function then calculates the areas of intersection of farms with each land unit, creating a table of farm identification numbers associated with each land unit, and the area of each section of each farm that overlaps each land unit. This data can then be manipulated by an expression that converts TB incidence data in the cattle to a TB prevalence figure in possums. Currently the system is designed so that the cattle TB incidence data from the farm with the highest incidence of all farms that overlap a land unit is used to determine the prevalence of TB in possums using a mathematical function to convert cattle incidence to clinical and subclinical prevalence of TB in the possum population.

Control programme

The possum control programme is entered by digitising polygons that represent the boundary of each control area. The control boundary file is a polygon layer that defines the areas across which users would like to differentiate control measures. For example, users may want to undertake annual possum control in one area and bi-annual in another. Alternatively, users may want to apply possum control and vaccination in one area and possum control only in another area.

Modelling population and TB dynamics

The software has been designed so that it can use different models of the dynamics of possum populations and of TB within these populations. It includes a set of tables in which the land units have a particular possum population and TB prevalence. Any model can be applied such that it updates these parameters for each land unit in the table.

We have run the model using the output from PossPOP, by applying the outputs in a semi-deterministic way. We have also run the model with a simple deterministic model of possum population and TB dynamics. It is possible to run TB-Spread with other models such as that of Barlow (1993).

Juvenile possum movement

Movement of juvenile possums is included as a separate function in the model, using a Poisson distribution to determine the distance that a juvenile possum moves from its land unit of origin. The movement of juvenile possums may be modified by naturally occurring geographic barriers such as wide rivers and mountain ranges. The number of juveniles available for migration is a function of the deterministic model output.

Running TB-Spread

The screen for running the model once the initial parameters have been set is shown in Figure 39.
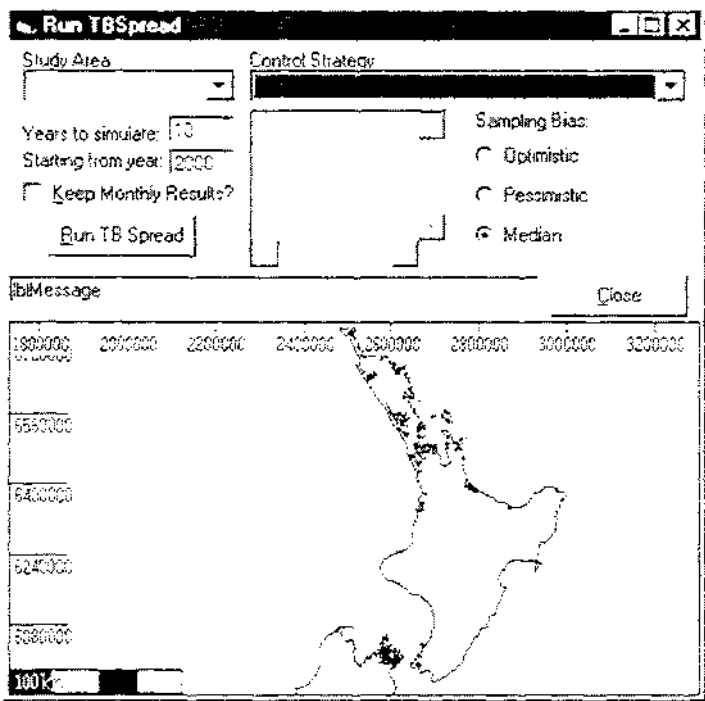

Figure 39. Screen for running TB-Spread once the initial parameters have been set.

Model output

Users have the option of printing out four maps that describe different features of land units for each time period over which the model was run. These are vegetation pattern classification, possum population, clinical prevalence of TB and subclinical prevalence of TB in possums.

Farm risk predictor module

The ability to classify farms within a region according to the risk of TB in the on-farm cattle or deer population being high, medium or low would enable TB managers to differentiate the intensity with which control measures are applied according to the risk of the farm having a TB problem. This is particularly useful in an area where the possum population has recently become infected with TB as farms at the greatest risk of having infected possums on their property could be targeted more intensively for surveillance and disease control activities.

We conducted research as a part of the larger study described in this thesis to identify geographic predictors of possum TB risk at the farm level (see Chapter 5). We identified a set of predictor variables. However, the nature of the variables makes them very specific for the particular geographic area in which the study was conducted. Further research is required to refine the model before coding of an operational module is undertaken in EpiMAN-TB. As an alternative to developing an expert system to predict farm risk, we have included in this software a function that enables farms to be displayed by cattle TB incidence or any other attribute included in the database.

Administration module

This module provides tools to import geographic data sets, and delete or rename existing data sets. Currently the system, has been designed to import raster data in the following three formats: Idrisi, EPIC (Image analysis software used by Landcare NZ), and Arc/Info text format for grid files. The other main feature of the administration module is the ability to define category sets. An appropriate category set can be chosen when importing data. This screen enables users to define display colours and labels for each category.

Conclusion

As current possum control strategies achieve the initial disease control objectives outlined in the Animal Health Board's five year strategy (Animal Health Board, 1995), strategies to eradicate TB from possum populations and prevent the recrudescence of TB will undoubtedly become more complex as the remaining problem areas have not responded to the standard approach already implemented. To ensure that cost effective strategies are developed to handle these situations, TB managers need access to all the information and information processing tools available.

EpiMAN-TB is a comprehensive piece of software that provides easy access to the information required for the major decisions that need to be made with respect to the management of TB-related possum control in an area. While the concept of EpiMAN-TB appears to be "high tech", the tool itself is designed to be very simple to use, providing an important means of giving field managers access to expert epidemiological information, spatial data processing tools and models to evaluate possum control strategies. We believe that use of this system will result in the development of more cost-effective possum control strategies both by means of providing important information and by promoting deliberation and analysis of the different options available.

Sprague (1993) wrote that the development of a DSS is an iterative process, with the most important four steps in the typical systems development process – analysis, design, construction, implementation – combined into a single step that is iteratively repeated. In this project we have completed these four steps for the administrative, hot spot predictor and evaluation of on-farm possum control strategies modules. The control strategy evaluation at the regional level is at the implementation stage. The farm risk module requires further research to refine a set of predictor variables that can be used as a rules base to construct an operational module.

Future plans involve the application of EpiMAN-TB in the field to develop a long term possum control strategy for a trial area that is located within the study area used in this project, as we already have geographical data sets for this area. A comparison of this programme with the existing possum control programme that is currently being implemented will enable us to identify areas where the additional information available within EpiMAN-TB resulted in the development of a control strategy that differed to the original. In addition, the process of implementing EpiMAN-TB in a field situation will enable us to identify how

the DSS helps the decision making process and also identify refinements that can be made to the system to better meet the needs of the decision makers using it. EpiMAN-TB has been designed with a flexible framework so that existing features can be modified and refined and additional features can be added as new information becomes available. This is a dynamic and flexible tool that will evolve as it is used in the field and managers identify additional information and/or functions that they would like to use.

Bibliography

Animal Health Board. (1995). National Tb Strategy: proposed national pest management strategy for bovine tuberculosis. Animal Health Board, Wellington, New Zealand.

Anon. (1986). History of Tb control scheme. In: S. Hennessy (Ed), Surveillance 13 (3): 4-8.

Anon. (1993). Wellington region operational plan for bovine tuberculosis, July 1993 - June 1998. MAF Quality Management, Masterton, New Zealand.

Barlow, N.D. (1991). A spatially aggregated disease/host model for bovine Tb in New Zealand possum populations. Journal of Applied Ecology 28: 777-93.

Barlow, N.D. (1993). A model for the spread of bovine Tb in New Zealand possum populations. Journal of Applied Ecology 30 (1): 156-64.

Barlow, N.D. (1995). Simple models for TB and possum control. In: F. Griffin, and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. Proceedings of the Second International Conference on *Mycobacterium bovis*. Otago Conference Series No. 3, University of Otago Press, Dunedin, New Zealand. 161-64.

Blanning, R.W. and King, D.R. (1993). Current research in decision support technology. IEEE Computer Society Press, Los Alamitos, California.

Clarke, G., and M. Clarke. 1997. Spatial decision support systems for health care planning. In S. Cropper, and P. Forte (Eds), Enhancing health services management. Open University Press, Buckingham, England. 43-70.

Cropper, S., and P. Forte. 1997. The nature of decision support systems. In S. Cropper, and P. Forte (Eds), Enhancing health services management. Open University Press, Buckingham, England. 19-35.

Efford, M.G. (1996) Simulating spatially distributed populations. In: R.B. Floyd, A.W. Sheppard, and P.J. De Barro (Eds), Frontiers of population ecology. CSIRO Publishing, Melbourne, Australia. 409-418.

Frysinger, S.P., Coppernan, D.A., and Levantino, J.P. (1996). Environmental decision support systems (EDSS): an open architecture integrating modeling and GIS. In: M. F. Goodchild, L. T. Steyaert, B. O. Parks, C. Johnston, D. Maidment, M. Crane, and S. Glendinning (Eds), GIS and environmental modeling: progress and research issues. GIS World Books, Fort Collins, Colorado. 357-61.

Hickling, G.J. and Efford, M.G. (1996). Assessing the risk of bovine tuberculosis infection for New Zealand cattle herds in TB vector risk areas (VRAs). Statistics in Ecology and Environmental Monitoring 2: 79-90.

Hutchison, J., DeLost, S., Liebhold, A.M., and Barrett, H.R. (1993). Coopers Rock Demonstration Project: a decision support system for gypsy moth managers. In: A. M. Liebhold, and H. R. Barrett (Eds), Spatial analysis and forest pest management. General Technical Report NE-175. Northeastern Forest Experiment Station, USDA Forest Service. Radnor, Pennsylvania, USA. 49-55.

Jackson, R. (1995). Transmission of tuberculosis (*Mycobacterium bovis*) by possums. Unpublished PhD thesis, Massey University, New Zealand.

Kessell, S.R. (1996). The integration of empirical modeling, dynamic process modeling, visulaization, and GIS for bushfire decision support in Australia. In: M. F. Goodchild, L. T. Steyaert, B. O. Parks, C. Johnston, D. Maidment, M. Crane, and S. Glendinning (Eds), GIS and environmental modeling: progress and research issues. GIS World Books, Fort Collins, Colorado. 367-71.

Livingstone, P.G. (1997). Update on the New Zealand TB situation. In: Proceedings of a Seminar on Possum and Mustelid Control Research. National Possum Control Agencies, Wellington, New Zealand. 17-30.

McGlinchy, A. (1996). Decision support tools for possum management. In: Improving conventional control of possums. The Royal Society of New Zealand miscellaneous series, volume 35. 35-38.

McKenzie, J.S. and Dymond, J.R. (in preparation). Automatic classification of a multispectral SPOT image to produce a vegetation map for modeling possum habitat in New Zealand. Geocarto International.

McKenzie, J.S. and Morris, R.S. (1995). EpiMAN-TB: a decision support system for the control of bovine tuberculosis. Surveillance 22 (4): 26-28.

McKenzie, J.S., Morris, R.S., Tutty, C.J., and Pfeiffer, D.U. (1997). EpiMAN-TB, a decision support system using spatial information for the management of tuberculosis in cattle and deer in New Zealand. In: R. T. Pascoe (Ed), Proceedings of the Second Annual Conference on Geocomputation. Spatial Information Research Centre, University of Otago, Dunedin, New Zealand. 271-77.

Morris, R.S. and Pfeiffer, D.U. (1995). Directions and issues in bovine tuberculosis epidemiology and control in New Zealand. New Zealand Veterinary Journal 43 (7): 256-65.

Morris, R.S., Sanson, R.L., McKenzie, J.S., and Marsh, W.E. (1993). Decision support systems in animal health. In: M. Thrushfield (Ed), Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine. 188-99.

O'Neil, B.D. and Pharo, H.J. (1995). The control of bovine tuberculosis in New Zealand. New Zealand Veterinary Journal 43 (7): 249-55.

Pfeiffer, D.U., Cochrane, T., Stern, M., and Morris, R.S. (1995a). In: Griffen F., and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. Proceedings of the Second International Conference on Mycobacterium bovis . Otago Conference Series 3, University of Otago Press, Dunedin, New Zealand. 165-67.

Pfeiffer, D.U., Cochrane, T., Stern, M., and Morris, R.S. (1995b). In: Griffen F., and G. de Lisle (Eds), Tuberculosis in Wildlife and Domestic Animals. Proceedings of the Second International Conference on Mycobacterium bovis . Otago Conference Series 3, University of Otago Press, Dunedin, New Zealand. 219-21.

Roberts, M.G. (1996). The dynamics of bovine tuberculosis in possum populations, and its eradication or control by culling or vaccination. Journal of Animal Ecology 65 : 451-64.

Sanson, R. L. (1993). Development of a decision support system for an animal disease emergency. Massey University.

Sanson, R.L.and Pearson, A.P. (1997). Agribase - A national spatial farm database. In: Proceedings of the 8th International Symposium on Veterinary Epidemiology and Economics. 12.16.1 - 12.16.3.

Sprague, Jr. R. H. (1993). A framework for the development of decision support systems. In Jr. R. H. Sprague, and H. J. Watson (Eds), Decision support systems. Putting theory into practice. Prentice hall International, New Jersey, U.S.A. 3-28.

Stärk, K.D.C., Morris, R.S., and Sanson, R.L. (1996). EpiMAN-SF - development of a decision support system for managing a swine fever epidemic. In: Proceedings of the 14th International Pig Veterinary Society Congress, Bologna, Italy.

Twery, M.J., Elmes, G.A., Schaub, L.P., Foster, M.A., Saunders, M.C., Liebhold, A.M., and Barrett, H.R. (1993). GypsES: a decision support system for gypsy moth management. In: A. M. Liebhold, and H. R. Barrett (Eds), Spatial Analysis and Forest Pest Management. General Technical Report NE-175. Northeastern Forest Experiment Station, USDA Forest Service. Radnor, Pennsylvania, USA. 56-64.

Watson, H. J., and R. H. Sprague Jr. (1993). The components of an architecture for DSS. In R. H. Spragure Jr., and H. J. Watson (Eds), Decision support systems. Putting theory into practice. Prentice Hall International, New Jersey, U.S.A. 99-110.

Williams, S.B. (1993) INFORMS-TX overview. In: A. M. Liebhold, and H. R. Barrett (Eds), Spatial Analysis and Forest Pest Management. General Technical Report NE-175. Northeastern Forest Experiment Station, USDA Forest Service. Radnor, Pennsylvania, USA. 85-92.

GENERAL DISCUSSION

Habitat predictors of possum TB risk at the possum home range level

In this study we have identified specific habitat and topographic factors associated with the probability of multiple and of single tuberculous possums being captured at a particular location (Chapter 2). Multiple TB possums were more likely to be caught on flatter land with a higher abundance of open dens (quality 1 dens) in large trees or tree ferns, or on flatter land with multiple fully enclosed dens (quality 4 dens). Single TB possums were more likely to be caught on flatter land with taller average tree height, a lower percent coverage of vegetation in the height range 0.3 – 2 metres, and a higher percent coverage of vegetation at the ground level compared to TB-negative possums. Considering these results in the light of other research findings, we believe that the most likely explanation of the location of possum TB clusters is that they are associated with particular patches of habitat that support a high number of favourable possum dens leading to local crowding of possums within these denning areas. This in turn may lead to a higher contact rate amongst possums, particularly during the process of finding a den each morning, and greater opportunity for the transmission of TB between possums. The opportunities for transmission may be even higher if there are one or two extremely favourable dens present as there may be increased competition amongst possums for such dens, leading to a higher rate of agonistic encounters around the entrance to these dens, close physical proximity, and/or a higher rate of den sharing, all of which provide suitable opportunity for the transmission of TB. The location of dens relative to ground level may also influence the degree of contact of possums around the denning area. For example, in areas where possums predominantly den above ground in trees there may be less direct contact between possums than in areas where possums predominantly den at or below ground level.

Previous studies on den factors that may be associated with the transmission of TB have focussed on identifying the frequency of simultaneous den sharing (Cole, 1993; Caley et al., 1998) and sequential den sharing by possums (Paterson et al., 1995). Observations of possums within dens during the day do not provide information on the contact that may have occurred between possums during the process of obtaining the den. We suggest that an appropriate direction for future research into factors associated with clustering of TB in possums relates to studying the nature of contacts amongst possums in the denning area, although this has proved difficult to achieve in earlier attempts.

A comparison of the ROC curves of the logistic regression models for distinguishing multiple and single TB sites from negative TB sites (see Chapter 2) shows that the habitat factors included in the models discriminate multiple TB sites more accurately from negative TB sites than they do single TB sites from negative TB sites. If we consider that the multiple TB site classification used in this study represents persistent possum TB clusters and the single TB site classification represents sporadic TB clusters, the results of this study suggest that habitat factors are more significant in the maintenance of persistent TB clusters than sporadic clusters. As described earlier, this could be explained by the effect of habitat on the probability of transmission of TB between possums, with TB clusters being persistent at locations where habitat favours transmission of TB, leading to the maintenance of TB at the location over a long period of time. Conversely, the habitat at sporadic sites is less favourable for the transmission of TB, hence the disease does not persist at these locations over the long term. In reality the locations of TB possums will fall along a gradient of highly persistent to less persistent. Persistent TB clusters are likely to have a very important influence on efforts to eradicate TB from possum populations, as TB possums are more likely to remain at the location of persistent clusters following population control and act as a focus from which infection can rebuild in the population. The ability to predict the locations of persistent TB clusters using habitat information facilitates the differential targetting of possum control resources at various habitat based on an assessment of possum TB risk with a consequent improvement in the cost-effectiveness of possum control programmes.

Application of GIS to TB management and research

Given the association of possum populations and of possum TB clusters with explicit spatial factors such as habitat types, geographical information systems (GIS) are a useful tool for the management and analysis of possum-associated data. A significant part of the studies presented in this thesis involved the investigation of available digital geographical data sets (Chapter 4) and of ways in which these data sets could be combined and analysed to represent habitat patterns that reflected factors associated with the distribution of possums and of TB possums (Chapter 5).

Vegetation coverage

Vegetation coverage is a fundamental data set for possum research and management as it has the greatest influence on the distribution of possums by means of its influence on food

sources and den sites. The digital vegetation data set that is currently available for a large part of New Zealand is the Land Cover Database (LCDB) which has been derived by manual interpretation of SPOT3 multispectral satellite images (Terralink, Wellington, New Zealand). We found that this data set did not meet our research needs as vegetation was classified into one of four broad categories (pasture, scrub, indigenous forest and exotic forest), and entire patches of habitat were generally classified into the one class. Thus much of the floristic heterogeneity of habitat patches had been lost in the classification process, and areas of cover less than one hectare were not included. As a part of this study we produced a vegetation data set for an area in the Wairarapa using an automated classification of a SPOT3 multispectral image (Chapter 4). We were interested in investigating the degree of accuracy that could be achieved with an automated classification procedure as this would be a more efficient way to produce maps of vegetation data for large areas of the country compared to manual classification. The final vegetation map comprised 8 classes and had considerably more detail than the LCDB, which enabled us to obtain more information on landscape patterns such as heterogeneity of patches of habitat. The accuracy of the classification was sufficiently high for our purposes. However, some of the classes were relatively broad and included a wide range of plant species. This resulted in different plant species that had different associated risks of TB possums being classified into the same vegetation class, which reduced the specificity of a TB risk classification system that used these vegetation classes. The remote sensing industry is rapidly expanding at present with the development of imagery with higher spatial and spectral resolution. We recommend further investigation of new imagery is required to identify that with a level of detail sufficient to be useful for possum control purposes.

Farm boundary data

As digital geographical data sets become available we can move from using static paper-based maps to dynamic computer-based maps with the flexibility to manipulate and present data in a multitude of different ways. The development of Agribase, a national digital data set of farm boundaries in New Zealand, is a valuable resource for TB management and research as farms are the basic geographical unit used in the administration and implementation of the TB management programme in New Zealand. Representing farms by their boundaries provides more information than representing farms by a single point. It enables users to visualise the extent of individual farms, to accurately identify the contiguity of individual farms with other farms and with other topographic features such as rivers, forests, lakes and roads. A digital

database of farm boundaries enables current maps of the distribution of different measures of TB in livestock to be easily drawn, enabling spatial patterns of the distribution of TB at the farm level to be identified (Chapter 3). This makes it easier to incorporate spatial information into decisions such as planning priority areas for possum control and for monitoring the effectiveness of control programmes. Hickling and Efford (1996) showed that the results of possum control programmes on farms varied with incidence of TB in cattle prior to control. These results suggest that it would be useful to group contiguous farms into areas according to TB incidence patterns and tailor programmes to the conditions depending on the specific disease management goals set for each area; for example, reduction of reactors, reduction of farms under quarantine, or prevention of re-infection. Digital farm boundary data makes it possible to use spatial analytical techniques to group farms in this way on the basis of contiguity and TB incidence patterns. Digital farm boundary data also enables maps of the distribution of the livestock population at risk per unit area to be drawn, which gives an indication of the sampling pressure for TB in possums.

As we obtain more epidemiological information on geographical risk factors associated with the distribution of TB possums at the farm level, digital farm boundary data enables these factors to be used to predict farms with a higher risk using computer technology. Such information is useful to target disease surveillance and vector control resources at higher risk farms which is particularly relevant in the TB-free areas surrounding VRAs to prevent the outward spread of VRAs.

Habitat and geographical predictors of possum TB risk at the farm level

In this study we used the percent years during which at least one TB-positive cattle beast was detected on a beef breeding farm as an indicator of the risk of TB possums being present on a farm in any one year (Chapter 5). We identified a group of habitat and geographical factors that could be used as predictors of farms with a higher risk of tuberculous possums being present on the farm. Farm size, total area of Pine and total area of Manuka/pasture were positively associated with percent TB-positive years, while distance from the coastal forest and proportion of the farm covered with Podocarp/broadleaved species were negatively associated. The model was strongly dominated by the variable 'distance from the coastal forest', which tended to swamp other variables that represented more detailed habitat patterns that may be associated with the risk of TB possums. This resulted in the model being very specific for the geographical area studied. By excluding 'distance from the coastal forest'

from the analyses we produced models of habitat variables which had a slightly lower predictive value, but which provided greater insight into the habitat factors that were associated with the odds of a TB-positive year occurring on a farm. Farms on which areas of habitat comprised a heterogeneous mix of vegetation classes, in particular Pine, Manuka/gorse and Manuka/pasture, interspersed in a patchy mosaic were associated with a higher odds of a TB-positive year occurring compared with farms on which the areas of habitat were more homogeneous. We believe that the most likely explanation for this effect is that the more heterogeneous the habitat the greater the chance of getting at least one location with the appropriate mix of conditions that supports a persistent TB hot spot. Farms with more heterogeneous areas of cover are likely to have both a higher number of persistent possum TB hot spots and a higher density of possums on the farm and surrounding area. Given the temporal variation of TB prevalence within possum TB hot spots a higher number of hot spots increases the probability that cattle will be exposed to at least one infectious possum within a year. TB infection may be more persistent at hot spots on these farms due to a combination of habitat factors that favour the transmission of TB between possums within hot spots, and a higher density of possums in the surrounding area which provides a large pool of susceptible possums and a pool of infected juveniles which disperse.

This study gave us valuable experience in analysing habitat and geographical patterns associated with the risk of TB possums occurring on a farm. As we have access to a more complete Agribase data set and more detailed vegetation information the methods applied in this study can be used to run further analyses and refine the model for predicting possum TB risk at the farm level.

Spatial patterns of TB at the farm level

We demonstrated the extent of significant spatial clustering of both the five-year cumulative incidence (1986-90) and percent TB-positive years on farms in an area of the Wairarapa (Chapter 3) using a spatial scan statistic which is available in the cluster analysis tool, SaTScan (Kulldorff et al., 1997). SaTScan both identifies if there is significant spatial clustering in the data set or not, and also identifies the units included in each cluster so that they can be mapped in a GIS. Visual display of the location of clusters provides greater insight into the nature of the disease and helps develop hypotheses regarding associated risk factors. By combining the clusters from both these TB measures we were able to identify different patterns of TB on farms in different geographical locations. A group of farms in the

north-western part of the study area had both a low cumulative incidence for the five-year period 1986-90 and a low percent of years during which TB-positive cattle were detected. A group of farms in the centre of the study area had a high percent of years during which at least one TB-positive cattle beast was detected but a low cumulative incidence, indicating that relatively low numbers of TB-positive cattle were detected during the five-year period 1986-90. A group of farms in the east of the study area had both a high percent of TB-positive years and a high cumulative incidence rate during the five-year period 1986-90.

Using a space-time scan statistic we showed evidence of space-time clustering of annual cumulative incidence of TB in cattle on only seven small groups of contiguous farms (2-5 farms) across a time period of one-to-two years, during the period 1979-95. This suggested that the temporal factors influencing annual cumulative incidence on an individual farm generally did not occur on neighbouring farms within the same year.

Through the course of this study we also identified how SaTScan could be applied as a useful surveillance tool for TB management purposes. The software identifies the location of groups of farms that have either a higher or a lower rate of TB incidence. Such clusters could form the basis of possum control strategies. For example, high incidence rate clusters can be targetted for disease reduction moving towards eradication, whereas low incidence rate clusters can be targetted for disease eradication and for maintaining freedom from disease. This cluster analysis tool can also be used to identify the location of individual farms that have a higher incidence than other farms in the surrounding area studied which would be useful in identifying problem farms that require extra disease control efforts. In areas of New Zealand where possum populations are believed not to be infected with TB (referred to as Surveillance Areas), analysis for spatial and space-time clusters could be used as an early warning system to indicate that infection has entered the possum population. Any spatio-temoral clustering of TB incidence involving multiple farms in Surveillance Areas would be considered an indication that a wildlife vector, most commonly the possum, may be associated with infection, and such cases would warrant thorough investigation.

EpiMAN(TB)

Decision making in relation to possum control for TB management purposes is very complex, involving data from multiple sources, including complicated geographical data. We have developed EpiMAN-TB as a tool to incorporate much of the epidemiological information that has resulted from this study, together with map display tools to facilitate the application of

this information and technology at the field level. The hot spot module classifies pixels within a raster image into one of three possum TB hot spot risk categories using a set of rules that combine categorical vegetation and slope data. As the particular combination of vegetation and slope categories that are associated with a higher risk of TB possums may vary between geographic regions, this module has been developed with the flexibility to allow the user to define different sets of rules by allocating different TB hot spot risks to different combinations of vegetation and slope data. A map of the resulting hot spot risk classification may be drawn over a vegetation map together with farm boundaries, rivers and other topographic features enabling the user to visually identify areas of habitat with a higher risk. This information can be used to design a possum control programme that is tailored to the particular conditions in the area of interest.

PossPOP is a spatial simulation model of the spread of TB in a possum population at the individual farm scale. It is a useful tool to evaluate the relative effect of different possum control strategies on population size and the prevalence of TB in the population. The model uses a real vegetation map to populate the area of interest with possums and with possum dens, and it can also use the hot spot risk map to more realistically represent the clustering of TB in certain habitat types.

The hot spot and PossPOP modules have been evaluated by two AgriQuality NZ veterinarians in the Wairarapa who indicated they would find the system useful as long as they had confidence in the accuracy with which the system predicted the location of possum TB hot spots and in the accuracy with which the possum TB model, PossPOP, simulated the dynamics of possum populations and of TB in these populations.

TB-Spread has been developed as a framework that can facilitate the application of different possum TB models within a spatial context at the regional level, on the basis of 25-hectare land units. This model can be used to evaluate regional control programmes which include a mosaic of operational areas each with a different history of possum control.

The development of EpiMAN-TB is an evolving process involving application of the software, feedback and refinement. The advantage of such a system is not only the information that is available to assist with decision making, but also the fact that it promotes deliberation of a problem. Sprague (1993)noted that one of the benefits described during research into the use of DSS was that managers considered a wider range of alternatives than they would have without the DSS. This system has been designed with a flexible framework

so that existing features can be modified and refined and additional features can be added as new information becomes available and as users identify additional information and/or functions that they would like to use.

Possum density

The localised scale at which the association between density and possum TB clusters appears to occur is worthy of further discussion, as this has important implications for measuring the effectiveness of possum control operations. The effect of possum density on TB transmission that leads to the development of TB clusters, appears to occur at a very localised scale, somewhere in the order of an area of typical diameter of 20-40 metres. Average measures of reduction in possum density over large areas do not measure density at the scale at which it influences transmission of TB between possums. Such density estimates are useful for assessing the effectiveness of possum control operations in reducing the overall possum population, which has been shown to reduce the abundance of TB possums. However, they are not necessarily useful in identifying how successful the operations will be in eradicating TB.

Implications for control

To move closer to the AHB's goal of eradicating wildlife-derived TB from domestic stock in New Zealand, new strategies need to be developed that focus on identifying and managing likely areas of residual infection in possums, while at the same time maintaining low population levels in the most cost-effective way. To ensure that the most cost-effective programmes are being implemented it is important that full use is made of current information on the understanding of the epidemiology of TB, and that information management tools such as GIS and DSS are utilised to develop strategies and manage the implementation of programmes. These tools can be used to design customised programmes for different areas and different farms to:

- eradicate TB (reduce incidence of TB in cattle and deer towards eradication in some areas);

- prevent TB from returning to these areas;

- prevent the expansion of existing VRAs.

There is a danger that those involved with developing policy, funding and implementing TB-

related possum control programmes are lulled into complacency that the satisfying results currently being achieved will continue and that intensification of control effort is unnecessary. History has shown us that this is risky as the disease will rebuild slowly but relentlessly as soon as the pressure of possum control is removed. It is important to be planning strategies for this stage. There already appears to be some public questioning of the scale of resources currently being spent on possum control for TB management purposes, and it is imperative that organisations responsible for implementing possum control are able to show that they are using the most cost-effective approaches available for their areas.

Bibliography

Caley, P., Spencer, N.J., Cole, R.A., and Efford, M.G. (1998). The effect of manipulating population density on the probability of den-sharing among common brushtail possums, and the implications for transmission of bovine tuberculosis. New Zealand Veterinary Journal 25: 383-92.

Cole, R. (1993). Denning behaviour of female brushtail possums (*Trichosurus vulpecula*) at Pigeon Flat, Otago. University of Otago Wildlife Management Report Number 41, Dunedin, New Zealand.

Hickling, G.J. and Efford, M.G. (1996). Assessing the risk of bovine tuberculosis infection for New Zealand cattle herds in TB vector risk areas (VRAs). Statistics in Ecology and Environmental Monitoring 2 : 79-90.

Kulldorff, M., Rand, K., and Williams, KG. (1997). SaTScan, version 1.0, software for the spatial and space-time scan statistics. Bethesda, MD: National Cancer Institute [Available at: http://dcp.nci.nih.gov/bb/SaTScan.html] .

Paterson, B.M., Morris, R.S., Weston, J., and Cowan, P. (1995). Foraging and denning pattern of brushtail possum and their possible relationships to contact with cattle and tuberculosis transmission. New Zealand Veterinary Journal 43 (7): 281-88.

Sprague, Jr. R. H. 1993. A framework for the development of decision support systems. In: R. H. Sprague Jr, and H. J. Watson (Eds), Decision support systems. Putting theory into practice. Prentice Hall International, New Jersey, U.S.A. 3-28.

APPENDIX

Appendix 1: Forms for recording possum habitat data

POSSUM HABITAT ANALYSIS PLOT SHEET

Date: Farm name: Trap/poison site no:

Photograph numbers and description:

File name and location of GPS readings:

Comments about the site:

TRAP LOCATION AND PLOT DIAGRAM (number each quarter, and note slope & aspect)

VEGETATION DATA SHEET

Date: Farm: Site No:

Hgt (note species)	DBH	Species % Abundance & Distribution & Shrub density					Ground (%)
		T2	T3	T4	T5	T6	
		(12m +)	(12-5m)	(5-2m)	(2-0.3m)	(<30cm)	
1.							V
							M
							B
							R
							W
		Epiphytes					
2.							V
							M
							B
							R
							W
		Epiphytes					
3.							V
							M
							B
							R
							W
		Epiphytes					
4.							V
							M
							B
							R
							W
		Epiphytes					

Distribution: C: Continuous, I: Interrupted, P: Patchy, S: Scattered
Braun-Blanquet: 1: <1 (0.5), 2: 1-5 (3), 3: 6-25 (15.5), 4: 26-50 (38), 5: 51-75 (63), 6: 76-95 (85.5), 7: 96-100 (98)
Ground cover: V: vascular vegetation, M: moss, B: bare ground & leaf litter, R: roots, branches & hollow logs,
          W: water

DEN SITE DATA SHEET

Date: Farm Name: Trap/Poison site No:

Den Site	1 Rel. abun	Locn	Qual	2 Rel. abun	Locn	Qual	3 Rel. abun	Locn	Qual	4 Rel. abun	Locn	Qual
Under live roots												
Under dead roots												
Dead hollow trunks												
Live hollow trunks												
Hollow stumps <2m												
In hollow logs												
Under hollow logs												
Under piles of branches												
Overgrown tree base												
Overgrown fallen log/branch												
Burrows												
Tomos												
Gorse												
Flax												
Blackberry												
Toetoe												
Carex spp												
Raupo												
Dense ground fern												
Under/in rocks												
Hollows in grass												
Recesses in bank												
Tree fern crowns												
Fallen punga leaves												
Epiphytes												
Cabbage trees												
Totara branches												
Buildings												

Relative abundance: **A**: Abundant, **C**: Common, **F**: Frequent, **O**: Occasional, **R**: Rare

Location: **A**: Above ground, **G**: Ground level, **B**: Below ground

Den quality:

1   Very open to the environment, direct sunlight, rain and wind. No structure defining a refuge site, e.g. epiphytes, in foliage of trees.

2   Open to direct sunlight, rain and wind. Some degree of protection offered by a 'structure', e.g. hole in the side of a tree trunk in which animals can sit.

3   Mostly protected but still allowing daylight to enter. Protected from the wind, rain and sunlight in all but extreme conditions.

4   Totally protected, e.g. hollow logs and long burrows